CAPTURING THE RATE OF GENE EVOLUTION

AND

THE TIMESCALE OF MAMMALIAN EVOLUTIONARY HISTORY

Céline M. Poux

Cover: *Hemicentetes semispinosus* (Ranomafana, Madagascar),
Picture taken by Miguel Vences

Printed by Ipskamp drukkers.

ISBN-13: 978-90-9024780-9

This work has been supported by NWO under the project number 050.50.214

Capturing the rate of gene evolution
and
the timescale of mammalian evolutionary history

Een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 7 december 2009
om 13.30 uur precies

door

Céline M. Poux
geboren op 26 Juni 1976
te Issy-Les-Moulineaux (Frankrijk).

Promotor: Prof. dr. W.W. de Jong

Manuscriptcommissie:
Prof. dr. E.W. Roubos
Prof. dr. S.E. Wendelaar Bonga
Prof. dr. J.A.M. Leunissen

Capturing the rate of gene evolution
and
the timescale of mammalian evolutionary history

A scientific essay in the
Natural Sciences, Mathematics and Computer Science

Doctoral thesis

To obtain the degree of doctor
from the Radboud University Nijmegen
on the authority of the rector prof. mr. S.C.J.J. Kortmann,
according to the decision of the Council of Deans
to be defended in public on Monday, 7 December 2009
at 10.30 a.m. precisely

by

Céline M. Poux
born on 26th of June 1976
in Issy-Les-Moulineaux (France).

Promotor: Prof. dr. W.W. de Jong

Manuscriptcommissie:
Prof. dr. E.W. Roubos
Prof. dr. S.E. Wendelaar Bonga
Prof. dr. J.A.M. Leunissen

Table of contents

P 11 **Chapter 1**
General introduction

P 73 **Chapter 2**
Sequence gaps join mice and men: phylogenetic evidence from deletions in two proteins

P 81 **Chapter 3**
Primate phylogeny, evolutionary rate variations and divergence dates: a contribution from the nuclear gene IRBP

P 115 **Chapter 4**
Arrival and diversification of caviomorph rodents and platyrrhine primates in South America

P 155 **Chapter 5**
Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes

P 179 **Chapter 6**
Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses

P 209 **Chapter 7**
General discussion and prospects

P 223 **Chapter 8**
Summary / Samenvatting / Résumé

P 237 **Curriculum vitae**

P 239 **List of publications**

P 241 **Acknowledgments**

En peinture

J'aime mon travail et le fais avec soin.
Mais la lenteur de l'ouvrage aujourd'hui me désespère.
Le jour me pèse. Les contours
ne font que s'assombrir. Il ne cesse de venter et de pleuvoir.
J'ai plus envie de voir que de dire.
Sur cette peinture, je regarde à présent
Un bel adolescent qui, près de la fontaine,
s'est allongé, essoufflé sans doute d'avoir couru.
Comme il est beau; quel divin midi à dû
S'emparer de lui pour l'assoupir. —
Je reste longtemps à le contempler de la sorte.
Et l'art, une fois de plus, me rend les forces qu'il m'a prises.

Ζωγραφισμένα

Την εργασία μου την προσέχω και την αγαπώ.
Μα της συνθέσεως μ' αποθαρρύνει σήμερα η βραδύτης.
Η μέρα μ' επηρέασε. Η μορφή της
όλο και σκοτεινιάζει. Όλο φυσά και βρέχει.
Πιότερο επιθυμώ να δω παρά να πω.
Στη ζωγραφιάν αυτή κυττάζω τώρα
ένα ωραίο αγόρι που σιμά στη βρύσι
επλάγιασεν, αφού θ' απέκαμε να τρέχει.
Τι ωραίο παιδί· τι θείο μεσημέρι το έχει
παρμένο πια για να το αποκοιμίσει. —
Κάθομαι και κυττάζω έτσι πολλήν ώρα.
Και μες στην τέχνη πάλι, ξεκουράζομαι απ' την δούλεψή της.

Pictured

My work, I'm very careful about it, and I love it.
But today I'm discouraged by how slowly it's going.
The day has affected my mood.
It gets darker and darker. Endless wind and rain.
I'm more in the mood for looking that for writing.
In this picture, I'm now gazing at a handsome boy
who is lying down close to a spring,
exhausted from running.
What a handsome boy; what a heavenly noon
has caught him up in sleep.
I sit and gaze like this for a long time,
recovering through art from the effort of creating it.

K.P. Kavafis / Κ.Π. Καβάφης

À ma petite grecque, patiente, telle
Pénélope attendant le retour d'Ulysse.

À ma famille, mon port d'attache…

Chapter 1

General Introduction

Trying to assess the rate of gene evolution has been an intensive field of research during the last decade and led to a rapid evolution of the available methods for dating speciation events on basis of DNA sequence comparisons. Even though important improvements have been made, debates are still going on concerning the reliability of the molecular age estimations made so far (e.g. Graur & Martin 2004; Hedges & Kumar 2004; Pulquério & Nichols 2007). Indeed, the amount and the nature of all parameters that must be taken into account to calculate datings is so extensive that it seems that even the best performing methods might lead to wrong estimations (Pulquério & Nichols 2007). How wrong these estimations can be is still not clear. In this introduction the first section will give an overview of the main components of the eukaryotic genome that can potentially be used in evolutionary studies. The following sections will review the main causes of the molecular evolutionary rate variability and their consequences, the state of the art of the dating methods and the role of fossil information for molecular datings. Finally, as a practical example, biogeographical patterns of colonization and speciation on islands will be described. It will be followed by a short overview of the chapters present in this thesis.

THE MAIN COMPONENTS OF THE EUKARYOTIC GENOME AND THEIR MOLECULAR EVOLUTIONARY RATES

The eukaryotic genome is made up of various segments of DNA of which the functions and rates of evolution greatly vary. There are two main components of the genome: the protein-coding genes, present in both nuclear and mitochondrial DNA, and the rest, called before "junk" or "selfish" DNA, as it was thought not to have any function. The latter include introns, retrotransposons (e.g. Long Interspersed Nuclear Elements or LINES, Short Interspersed Nuclear Elements or SINES), DNA transposons, simple sequence repeats, segmental duplications, etc. (Fig.1). Their real function has not yet been completely deciphered but they are involved in many processes, including transcriptional and posttranscriptional regulation, chromosome replication, genomic imprinting, RNA processing, modification and alternative splicing, mRNA stability and translation, and even protein degradation and translocation (reviewed in Shabalina & Spiridonov 2004).

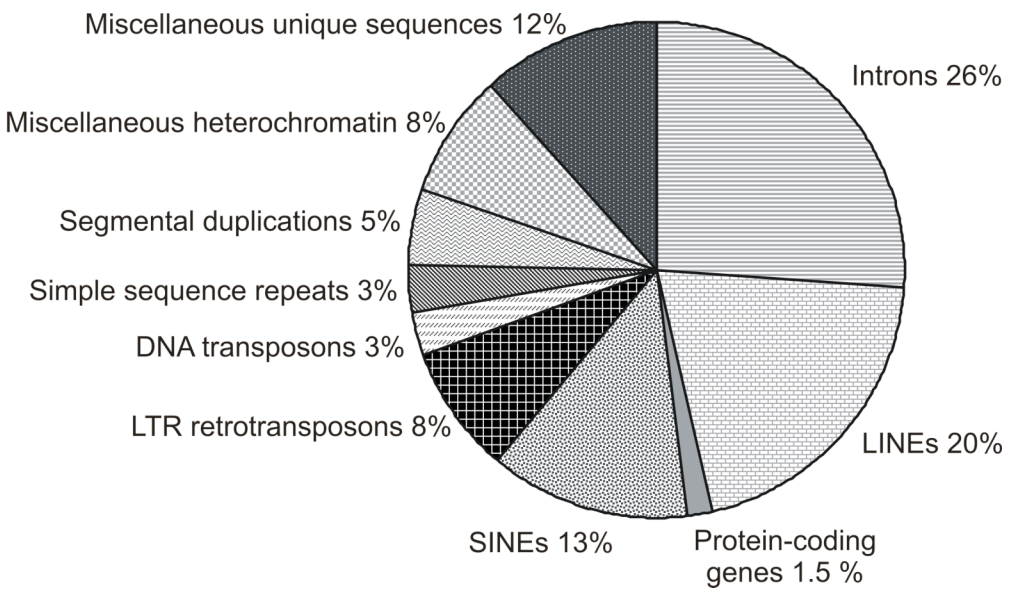

Figure 1: **The main components of the human genome.** LINES: Long Interspersed Nuclear Elements, SINES: Short Interspersed Nuclear Elements. Modified from Gregory (2005).

Most prokaryotic chromosomes consist almost entirely of protein-coding genes (Gregory & DeSalle 2005); in contrast, the size of eukaryotic genomes is not correlated with the number of genes they contain or the complexity of the organisms that carry them. The protein-coding genes represent a small fraction of most eukaryotic genomes; the human genome might contain as few as 20,000 to 25,000 genes (Lander et al. 2001)**,** comprising less than 1.5% of the total genome sequence (IHGSC 2001, 2004; Claverie 2005).

The majority of the genome is made up of non-coding DNA that represents 98.5% of the human genome. Shortly after the discovery of non-coding sequences within coding genes (introns) they were suggested to account for the pronounced discrepancy between gene number and genome size (Gilbert 1978). However, eventually introns were found to account only for a little more than a quarter of the draft human sequence (IHGSC 2001). Pseudogenes (non-functional copies of coding genes at the origin of the term 'junk DNA' [Ohno 1972]) even when taken altogether — 'classical pseudogenes' (direct DNA to DNA duplicates), 'processed pseudogenes' (copies that are reverse-transcribed back into the genome from RNA and therefore lack introns) and 'Numts' (nuclear pseudogenes of mitochondrial origin) — also comprise only a relatively small portion of the mammalian genomes. The human genome is estimated to contain about 20,000 pseudogenes (Torrents et al. 2003).

Conversely, transposable elements (Table 1) make up a large portion of the eukaryotic genome, with some elements (for example, the *Alu* element) present in more than a million copies in human. They are divided into two general classes according to their mode of transposition. Class I elements transpose through an RNA intermediate. This class comprises LINES, endogenous retroviruses, SINES (e.g. *Alu* elements) and Long Terminal Repeat (LTR) retrotransposons. Class II elements transpose directly from DNA to DNA, and include DNA transposons and Miniature Inverted Repeat Transposable elements (MITEs).

Table 1. Characteristics of various Transposable Elements.

		Type	Size	Autonomous[b]
Class I Retrotransposons	LTR[a] retrotransposon	Viral	100bp to 5 kb	yes
	Non-LTR retrotransposons	SINES	< 500 bp	no
		LINES	1-7 kb	yes
Class II DNA transposons	DNA transposons (MITEs excluded)		1-25 kb	yes and no
	MITEs		100-600 bp	no

[a] Long Terminal Repeats flanking functional genes, used to integrate into host genome.
[b] Retrotransposons are autonomous if they possess a reverse transcriptase activity, and DNA transposons if they possess a Transposase.

Perfect or near-perfect tandem iterations of short sequence motifs (called simple sequence repeats or microsatellites) are extremely common in eukaryotic genomes, and in the case of the human genome they are found at hundreds of thousands of places along chromosomes. Microsatellites are among the most variable types of DNA sequence in the genome (Ohno 1972) and in contrast to unique DNA, microsatellite polymorphisms derive mainly from variability in length rather than in the primary sequence. Due to their fast rate of evolution microsatellites are not used in phylogenetic studies but rather in population genetics.

In addition to all these genome components (but microsatellites) that can be used in phylogenetic analyses, rare genomic changes (RGCs) such as indels (insertions/deletions within a given DNA sequence), retroposon integrations, mitochondrial and chloroplast gene order changes, gene duplications and genetic code changes, provide a suite of complementary markers with enormous potential for molecular systematics (Rokas & Holland 2000): they are rare and therefore unlikely to occur in parallel or to reverse, i.e. considered as poorly homoplastic. This means that it is extremely improbable (even though it happens) to find two or more individuals with a comparable RGC not derived from a common ancestor.

Consequently the RGC can help to define a reliable frame on which divergence ages can be estimated. However, they cannot be used in the process of dating itself.

Because all these components of the genome have distinctive or no functions, they are under highly different selective pressure and therefore their overall rate of evolution greatly varies. These differences allow choosing the molecular marker best suited to answer a particular question. For example, fast-evolving mitochondrial genomes and hypermutable nuclear DNA are used to construct rapidly ticking clocks for the fine resolution of more recent events over relatively short timescales, including the evolution of populations and closely related species. Slowly ticking clocks, which are based on nuclear coding DNA, are used to time deep divergences, whereas very highly conserved proteins are used to establish the timing of the earliest divergences in the tree of life. Therefore, the choice of a molecular marker is a crucial issue as its particular rate of evolution must be adequate for the specific purpose of a study, meaning that the marker should be variable enough without being homoplastic.

THE VERSATILE MOLECULAR CLOCK

The molecular-clock hypothesis

The molecular-clock hypothesis was put forward in the early 1960s, only a few years after DNA was established as the hereditary material and the first protein (insulin) was sequenced (Sanger 1959). The molecular-clock hypothesis represents the idea that molecular evolution proceeds at an approximately uniform rate over time, which should therefore be similar among morphologically diverse species with vastly different life history traits. Because of these underlying assumptions, the molecular-clock hypothesis led to many controversies since its initial presentation.

The use of molecular clocks began with the analysis of mammalian protein sequences. In 1962, Zuckerkandl & Pauling observed that the number of amino acid differences between the hemoglobins of different species was correlated with the time passed since they diverged. The same result was observed for other proteins, showing a constant rate of molecular evolution across species, each protein with a characteristic rate of amino acid substitution (Margoliash 1963; Doolittle & Blombaeck 1964). This relative constancy of evolutionary rate for a given protein was

unexpected; it had been assumed, as with morphological evolution, that there would be large variations in the rate of change, both between species and over evolutionary time. The relatively constant rate of protein evolution could be explained both from a 'selectionist' and a 'neutralist' point of view. According to the selectionist theory, Margoliash & Smith (1965) proposed that the rate of evolution increases owing to positive selection during short time intervals, producing similarity when rates are averaged over long time spans. This theory considers that most amino acid changes are either positively or counter-selected (Fig.2a). By contrast, Zuckerkandl & Pauling (1962, 1965) argued that most observed substitutions are selectively neutral and have been fixed by random genetic drift in the population. This latter assumption was the foundation for the neutral theory of molecular evolution (Kimura 1968, 1983).

Kimura (1968, 1983) explained the constant rate of protein evolution by assuming that most amino acid substitutions are effectively neutral, namely that they have little effect on the fitness of the individual carrying them. By consequence, the rate of amino acid change would only be very slightly affected by natural selection, as advantageous mutations are rare and deleterious ones would disappear rapidly. For a haploid taxon, the number of mutations occurring per time period is equal to Ne (effective population size) * μ_0 (mutation rate), and the probability of fixation of a neutral mutation is 1/Ne, therefore the number of neutral mutations fixed per time period is $Ne*\mu_0/Ne= \mu_0$. As a consequence, in the neutral model, the effective population size does not influence the rate of molecular evolution. The neutral theory predicts therefore that, for a given mutation rate (μ_0) and proportion of neutral sites, the rate of molecular evolution should be constant, and the greater the proportion of neutral sites, the faster the rate of molecular evolution will be (Fig.2b) (Dickerson et al. 1971).

The neutral theory has been strongly criticized at that time (King & Jukes 1969), being called "non-Darwinian evolution" as its essential basis was the assumption that molecular evolution was independent of natural selection. Moreover, under the neutral theory, molecular and morphological evolution are dichotomous, i.e., the former occurs predominantly by random drift at an almost uniform rate, and the latter by natural selection depending on environmental conditions. Nevertheless, evolution of genes should obviously be responsible for morphological evolution.

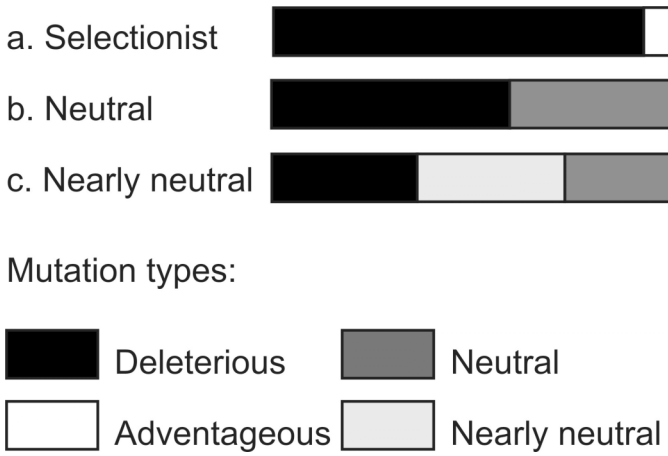

Figure 2: **Proportions of different types of mutations according to the selectionist, neutral and nearly neutral theories.** (a) The selectionist theory assumes that all mutation would affect fitness; they would be therefore advantageous or deleterious, but not neutral. (b) The neutral theory assumes that, for most proteins, neutral mutations exceed advantageous ones. Therefore the rate of evolution of such a gene would mainly depend on the proportion of neutral mutations. (c) In the nearly neutral theory, nearly neutral mutations will affect the rate of evolution depending on the size of the population carrying them. Modified from Bromham & Penny (2005).

The validity of the strict molecular clock was immediately questioned. One of the most important debates concerned the "hominoid slow down" (reviewed in Goodman 1985) that describes an overall slow down of the molecular substitution rate within the hominoid clade. When the information contained in DNA sequences became accessible, first by measuring the strength of heterologous DNA duplexes (DNA-DNA hybridization; Laird et al. 1969) and later on, in the beginning of the 1980's, by the development of direct-sequencing techniques, a discussion started about the large difference in substitution rate between rodents and primates (Wu & Li 1985; Easteal 1985; Britten 1986; Li 1993; Easteal et al. 1995). However, at that time it was difficult to assess and compare molecular evolutionary rates between species because of the influence of many unknown parameters on the analyses. For example, during this period the relative rate test (RRT) was widely used to test for variation in the rate of molecular evolution between lineages (Fitch 1976); this test compares the distance between each of two taxa and an outgroup to determine the relative amount of change in each lineage since their last common ancestor. However, if the application of this test is based on incorrect phylogenetic assumptions or if the rate of synonymous (neutral) transitions (K_s) is saturated (e.g. when the compared taxa are genetically too distant), the outcome of the test is biased. This explains why, depending on different studies, authors could generalize or not the existence of a global or universal clock as applicable to any organism.

However, thanks to the accumulation of experimental evidence, the idea emerged that molecular evolutionary rate differences between species were largely due to generation-time effects (Kohne 1970; Wu & Li 1985), meaning that a fast evolutionary rate was related to a high mutation rate in short generation time species. The variation in rates of molecular evolution was such that the "Neutral Theory" was not sufficient to satisfactorily explain why the clock looked as if it depended either on real (chronological) time, when it was measured with proteins, or on generation number when DNA was used (Ohta 2003). Ohta and Kimura (1971) and Ohta (1973, 1987, 2002) provided then an important extension of the Neutral Theory by recognizing the crucial role of the effective population size in the influence of selection (Fig. 2c). The fundamental principle of the 'Nearly Neutral Theory' is that slightly deleterious mutations will tend to be removed by selection in very large populations, but can be fixed by chance in smaller populations, in which selection can be overpowered by genetic drift. As a general inverse relationship between generation time and population size has been described, the nearly neutral theory predicts that non-synonymous substitutions will occur approximately evenly distributed over time because the effect of population size on substitution rates (faster substitution in smaller populations due to fixation of mildly deleterious mutations) could be cancelled out by the generation time effect (slower rates in long generation time organisms). However, this inverse relationship is far from being general. In any case this compensation between population size and generation time should be more effective at the protein than at the DNA level as the excess of neutral synonymous mutations in short generation organisms is only subject to drift and therefore no compensatory phenomenon can be applied. In conclusion, the nearly neutral theory leads us to expect that the rate of molecular evolution can vary in three ways depending on the mutation rate, the population size/generation time and the selection strength.

These theories as well as observations support the prediction that the degree of divergence between orthologous genes tends to increase with the time since separation (Fitch 1976). In practice, then, by counting the amount of difference between the DNA of two species, it is possible to place an evolutionary event back in time, if one knows the rate at which a given gene evolves or by using a paleontological timeframe to calculate it.

The molecular clock is a sloppy clock

The molecular-clock hypothesis proposes that molecular evolution of a given gene occurs at rates that persist through time and across lineages provided that the basal mutation rate remains unchanged. A notable feature of this hypothesis is that every component of an organism's genome has an independent clock, ticking at a different rate, but all of them are measuring the same events. The differences in their 'ticking' rates are due to the level of selection to which they are subjected.

However, deviations from rate constancy occur between lineages, as well as at different times along a given lineage, both factors having significant effects. Firstly, the 'tick' rate of the clock is sloppy, i.e., it ticks at highly irregular intervals that contribute to a 'residual effect' (unevenness of substitution rate in a lineage), and secondly, the changes in mutation rate lead to a 'lineage effect' (variation in substitution rate between lineages) (Ayala 1999; Bromham & Penny 2003). As a result of these effects the molecular clock can no longer be described as a Poisson distribution (Ohta & Kimura 1971), with equal mean and variance, but as an "overdispersed" Poisson distribution with a variance typically larger than the mean (Gillespie 1991; Li 1997; Bickel 2000; Cutler 2000). In addition, a recent study (Bedford & Hartl 2008) shows that substitution counts are better described by a negative binomial distribution rather than a Poisson distribution.

The "residual effect"

Can we know how these extra variations arise in order to predict the pattern of the deviations? Several biological variables could affect the number of substitutions per unit of time.

First there are the processes changing the balance between the relative influences of selection and drift, either for specific genes, for sites within genes or across the whole genome. Because of functional constraints, substitution rates vary among the positions of a protein, but are usually assumed to be constant at a given site during evolution. The distribution of the rates across the sequence positions generally fits a Γ distribution (Yang 1996), which describes the proportion of sites evolving at a given rate. However, it has been convincingly demonstrated that the substitution rate at a given position is not always constant throughout time. This phenomenon of within-site rate variations over time is called heterotachy (Lopez et al. 2002). For

example, via the solvent-accessibility environments and pairwise amino acid interactions, the three-dimensional protein structure has an important effect on the patterns of substitutions. The interdependence between amino acid sites lead to an autocorrelation between substitution rates, and therefore increase the variance in the number of substitutions per time unit compared to a simple Poisson process. As a consequence, when the three-dimensional structure of a protein evolves, possibly leading to shifts in function, the position and proportion of variable sites in the sequence change too. Such a process has been described by the covarion (or concomitantly variable codon) model of protein evolution, which allows the hypothesized rate of evolution of individual codons in a set of nucleotide sequences to vary in an autocorrelated manner (Fitch & Markowitz 1970; Galtier 2001; Penny et al. 2001).

After gene duplication the selection pressures applied to the two daughter genes may vary. The strength of selection applied on particular nucleotide sites can change, for instance producing either bursts of substitutions, when molecules get adapted to a new role (e.g. Zhang et al. 2002), or non-functional pseudogenes. In that last case the selection coefficient of the entire gene will change. Thereafter, any change in the non-functional sequence has no further effect on the fitness, as all substitutions are neutral. These pseudogenes have therefore a fast rate of molecular evolution.

As organisms are continually adapting to their physical and biotic environments, which change endlessly in patterns that are unpredictable and of different significance to different species, it is possible that the overall rate of adaptation or morphological change of a species might influence the molecular clock. However, so far there has been little evidence to support this proposition from either experimental studies (Papadopoulos et al. 1999) or comparative analyses (Bromham et al. 2002).

A second source of variability that could affect the number of substitutions per unit time comes from the effect of population size on the fixation rate of mutations. Small populations are more severely affected by stochastic fluctuations in allele frequencies, so genetic drift (the effect of random sampling error on allele frequencies) can overpower selection for alleles with small selection coefficients. Therefore, the fixation of nearly neutral alleles with small selective effects is expected to be highest in small populations (Ohta 1987; 2002). The effective population size of

a lineage can change over time; for example, if a population undergoes a marked reduction in size due to an environmental catastrophe, this event might be accompanied by a burst of fixation of nearly-neutral alleles. In this way, population size fluctuations might add to the sloppiness of the molecular clock.

These parameters render the rates of molecular evolution difficult to model. What remains of the molecular clock is that evolution is a time-dependent process, and thus the longer the time elapsed, the larger the number of changes. One might think that fluctuations in rate are scattered randomly over the phylogeny. If this is true, then local variations in substitution rate might average out over long time scales, and simply add noise to the molecular clock, making date estimates imprecise and confidence intervals very large. Therefore, it shows the importance of considering rate variability when using molecular clocks to estimate dates of divergence.

The "lineage effect"

In addition to imprecision arising from the variability in the substitution process, the tick rate of the clock can also vary consistently between different species, such that some branches of the phylogeny have a faster rate of molecular evolution than others. In this case, molecular date estimates could be systematically inaccurate (e.g. Bromham et al. 2002; Douzery et al. 2003). The nearly neutral theory of molecular evolution leads to the expectation that the rate of molecular evolution can vary in three ways: through changes in mutation rate, population size or selective coefficients. Because these factors can differ between species or over time, they give rise to lineage effects (variation in substitution rate between species) (Gillespie 1991):

Mutation rates clearly vary between taxa (Drake et al. 1998) and much of this variation is due to differences in DNA repair mechanisms. Mutations arise both through unrepaired errors during DNA replication and DNA damage. Both of these processes are governed by an array of enzymes that can vary in repair efficiency between species. Errors during DNA replication depend on the accuracy of the DNA polymerase. For example, within mammalian cells, mitochondrial DNA is copied by a DNA polymerase that has a higher error rate than other polymerases, contributing to the higher mutation rate of mitochondrial genes over nuclear genes. The extent of DNA damage is influenced by the life history of a species and, potentially, by environmental variables. If by-products of metabolic process (such as oxygen radicals) have a mutagenic effect on DNA (Rand 1994), then species with higher

metabolic rates might generate higher concentrations of mutagens and incur more DNA damage (Martin & Palumbi 1993).

Even species with the same basic DNA replication efficiency per cell division could vary in the rate of accumulation of DNA copy errors. For a given rate of copy error, a species that copies its DNA more frequently per unit time will accumulate more copy-error mutations than a species with a longer generation time during the same time interval. In this sense, the mutation rate might be best measured as a per-generation rate.

The nearly neutral theory also predicts that differences in effective population size between lineages could influence substitution rate for the same reasons already described previously. Given the lack of a relationship between phenotypic evolution and molecular evolutionary rates (Papadopoulos et al. 1999; Bromham et al. 2002; Wyles et al. 1983), it seems unlikely that the observed faster rate of molecular evolution in more rapidly diversifying lineages is a result of species adapting to new niches. However, it could be related to population size effects if rapidly speciating lineages undergo frequent population subdivision (Bromham & Cardillo 2003).

The "time dependency of molecular evolutionary rates"

There have been several inklings (Garcia-Moreno 2004; Howell et al. 2003; Lambert et al. 2002) that the rate of molecular evolution accelerates when measured over evolutionarily short timescales. Ho and colleagues (2005), analyzing primate and bird data sets, confirmed this trend and relate it to the action of selection and genetic drift (Ho et al. 2005; Penny 2005). Pedigree studies estimate the mutation rate (mutations occurring per gamete per generation), whereas long-term evolutionary comparisons estimate the substitution rate (mutations fixed per generation). For instance, deleterious mutations are not expected to become fixed in large populations; they can nevertheless persist in the population for long periods of time and the average time before their loss correlates with their deleteriousness. Thus, as observation times diminish, one should observe a greater proportion of slightly deleterious mutations that have yet to be lost, with the most deleterious ones observed only in the short-term pedigree studies. This produces the apparent acceleration in substitution rate as the separation times between sequences decrease. Ho et al. (2005; 2007) observed this effect in different data sets, and in each case it was found that the

relation between the age of the calibration and the rate of change could be described by a vertically translated exponential decay curve (Fig.3). According to these authors, the molecular rate decreases continuously from generations (that is, in pedigree studies) to local and then widespread populations. Finally, at about 1-2 million years, it would reach a plateau describing a long-term evolution of the molecular rate. If correct, these results carry major implications for molecular evolutionary biology (Penny 2005). Therefore, Bandelt et al. (2006) and Emerson (2007) assessed the results of Ho et al. (2005) critically. Even though Emerson (2007) agrees on the fact that pedigree data have a remarkably high mutation rate compared with the more moderate mutation rates of phylogenetic studies, he considers that the time scale upon which the pedigree rate converges to the evolutionary rate is very much shorter than the timescale Ho et al. (2005) have focused on (Fig.3), and it is debatable whether this convergence would follow an exponential distribution. The challenge of formulating a single theory that operates smoothly over disparate time scales still remains. The consequences of Ho and colleagues' conclusion are practical in the sense that many time estimates might require recalculation, including the times of events in recent human evolution.

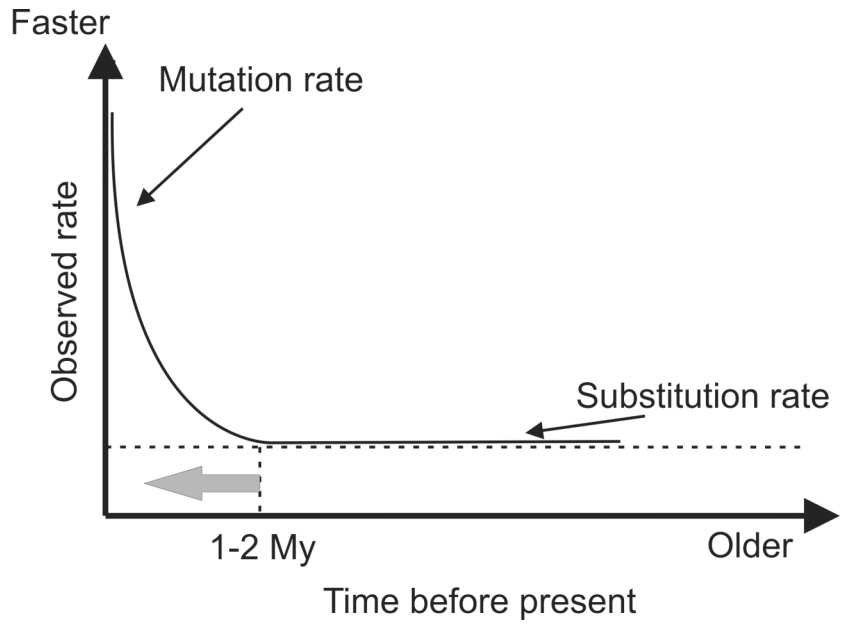

Figure 3: **A molecular rate curve showing the transition between the instantaneous mutation rate and the long-term substitution rate.** According to Ho et al. (2005), the transition period passes through a critical region around 1-2 My. Modified from Ho & Larson (2006).

The emphasis of current research (since the 1990's) has changed from testing the existence of a global DNA clock, which we know now doesn't hold, to quantifying the extent to which rate varies and determining the causes in order to increase the accuracy of age estimates.

Variations in molecular evolutionary rate influence phylogenetic reconstructions and datings

Heterogeneity of molecular evolutionary rates and phylogenetic reconstruction

There are several cases where standard phylogenetic reconstruction methods yield statistically well supported but wrong trees. These systematic errors are caused by the mutational saturation of the sequences: if some positions have undergone multiple substitutions, this will blur the phylogenetic signal, and thereby increase the probability for several species to display convergent sequence patterns (homoplasies) at those positions. Many reconstruction methods are unable to correctly identify these convergences, and will instead interpret them as shared derived characters. As a consequence, they will be misled towards reconstructing a wrong tree. A typical instance of this phenomenon, called the long-branch attraction (LBA) artifact (Felsenstein 1978), occurs when two phylogenetically distant species, evolving significantly more rapidly than the rest of the taxa (hence having long branches), deceivingly appear as closely related in the estimated tree. This artifact occurs mainly under poor taxon sampling (Blair et al. 2002; Rokas et al. 2003; Goremykin et al. 2003; Wolf et al. 2004). Similarly, when a distant outgroup is used, a divergent species may be attracted by the long branch separating the in- and the outgroup, and thus be artifactually put at a basal position (d'Erchia et al. 1996; Philippe & Laurent 1998).

Several alternative methods have been proposed to deal with LBA. They use a combination of a better taxon sampling — either by replacing fast-evolving taxa by slow-evolving close relatives (e.g. Soltis et al. 2004) or by 'breaking' a long branch by adding more taxa (Hendy & Penny 1989) — and a more careful selection of sites (Philippe et al. 2000) or sequences (Philippe et al. 2005). Altogether this makes it possible to converge to reliable phylogenies. And indeed, at various evolutionary scales (mammals [Murphy et al. 2001a; Murphy et al. 2007], metazoans [Philippe et

al. 2005], plants [Stefanovic et al. 2004], eukaryotes [Simpson & Roger 2004]), a consensus is gradually emerging regarding the overall evolutionary relationships in all those groups.

The above methods are exclusively focused on the quality of the data, leaving open the problem of understanding why current reconstruction methods (even the probabilistic ones) are so prone to systematic artifacts. This is explained by the lack of robustness of current models to capture the actual nucleotide evolution, under-estimating their true level of saturation. Many directions have already been explored to improve phylogenetic models, by accounting for compositional biases (Galtier & Gouy 1998; Foster 2004), across-site rate heterogeneities (Yang 1994, Felsenstein & Churchill 1996), substitution processes (e.g. Koshi & Goldstein 2001; Lartillot & Philippe 2004; Pagel & Meade 2004; Lartillot et al. 2007), or by acknowledging the variation of site-specific rates with time (heterotachy) (Tuffley & Steel 1998), the non-independence between sites (covarion models) (Galtier 2001; Robinson at al. 2003; Siepel & Haussler 2004; Rodrigue et al. 2005), etc. Some of these models have indeed resulted in improved phylogenetic inference (e.g. Philippe & Germot 2000; Brinkmann et al. 2005, Lartillot et al. 2007).

The heterogeneity of molecular evolutionary rates is one of the parameters that have to be taken into account when reconstructing a phylogenetic tree. The choice of the right model of base-pair evolution and of the right method of reconstruction is crucial, too. Building a reliable phylogenetic framework is the first step when doing molecular dating. Estimating ages on the wrong tree will necessarily lead to biased estimates; however, tree reconstruction is not our main interest in this thesis and therefore it will not be developed further in this introduction.

Heterogeneity of molecular evolutionary rates and molecular dating

Molecular evolutionary theory leads us to expect two types of error in molecular clock estimates. First, the sloppy nature of the substitution rate results in large variance around the amount of genetic difference expected for any given time period, adding a large degree of imprecision to molecular date estimates (Bromham et al. 1998). Second, the nearly neutral theory predicts that the rate of molecular evolution is influenced by mutation rate, population size and the relative proportions of sites with different selective coefficients; these factors differ between genes, between species and over time, potentially resulting in consistently over- or under-

estimated dates (Bromham et al. 2000a). Problems with the accuracy and precision of molecular clocks can be demonstrated by comparing studies that use different genes, calibrations or methods, and produce different date estimates for the same divergence (Bromham et al. 1999).

The variance in the clock, both due to the sloppiness of the tick rate and the lineage variation in rates, is assessed by confidence intervals accompanying the date estimates. Such confidence intervals allow molecular datings to be used to test evolutionary hypotheses within the limits of the accuracy and precision of molecular clocks, by asking whether the range of possible date estimates is consistent with a specific evolutionary hypothesis (Rambaut & Bromham 1998), and to point out cases for which molecular date estimates from different studies conflict significantly. These confidence intervals also take into account the uncertainty in the relationships between taxa (the evolutionary tree), errors in calibration points and uncertainties in the mechanism of evolution.

As described in recent articles (Ho et al. 2005, 2007), the debated "time dependency of molecular rates" might have been responsible for recurrent dating errors, and therefore provides a strong warning against extrapolating molecular rates across the population-species boundary, in either direction (extrapolation of phylogenetic rates to population-level data or the opposite, the extrapolation of pedigree rates to deeper timescales), unless the transition is well understood and has been quantified.

In conclusion, to estimate the timing of cladogenetic events in a valid manner some guidelines have to be followed, including but not limited to the following:

(i) Estimate dates within a relaxed clock framework that permits molecular rates to vary among branches (e.g. Thorne et al. 1998), enabling each branch to have its own rate. This only works if there is a reasonably dense taxon sampling.

(ii) Use suitable fossil calibration points, of which the ages are implemented in the dating programs as probability densities and not as hard bounds. Ideally, multiple calibration points should be used, spread over the tree as well as close to the nodes to be dated.

(iii) If necessary and feasible, use an accurately estimated molecular rate curve to derive the rate needed for the timescale in question (Ho & Larson 2006).

THE EVOLUTION OF DATING METHODS

Dating methods based on the molecular clock rely on accurate estimates of the genetic distance between sequences, but in some cases this measurement might be biased, as we have seen in the previous section. The challenge is to develop methods that enable this valuable source of historical information to be exploited, while recognizing their limits of precision.

All molecular dating methods convert measures of the genetic distance between sequences into estimates of the time at which the lineages diverged. Genetic distance estimates require topologies (the branching tree structure of the relevant lineages) and branch lengths (the inferred number of substitutions that have occurred in each lineage). Converting genetic distances into measures of time also requires one or more 'calibration points', which are externally derived dates, usually based on fossil or biogeographic evidence. To extrapolate from these 'known' dates to the rest of the tree, molecular dating has commonly relied on rate constancy, the assumption that molecular evolution occurred at a steady rate over the whole tree. Unfortunately, as we have seen in the previous section, there is increasing evidence that the assumption of rate constancy is often violated, and that the DNA of even closely related species might evolve at different rates (Bromham & Penny 2003). Here, we review methods for dealing with rate heterogeneity in molecular dating, and pay particular attention to those methods that incorporate multiple rates directly into the estimation.

Excluding anomalous sequences: the relative rate test (RRT)

A common and logical response to the problem of rate heterogeneity has been to try identifying lineages with 'anomalous' rates, or genes or sites that are most subject to rate variation, and then exclude these from the dating analysis. Many dating studies that assume rate constancy will be of this type (Hedges & Kumar 2003). To identify rate-variable lineages, various tests of rate constancy can be used (e.g. Sarich & Wilson 1973; Wu & Li 1985; Takezaki et al. 1995). The 'relative rate test' was introduced for protein sequences by Sarich & Wilson in 1973; all other developed methods are variants thereof. The RRT requires that we know the sequence of a

protein in three species, A, B and C, for which we also know their phylogenetic branching order. Considering C as outgroup, we can infer the amounts of change in the two lineages from the common ancestor of A and B to the modern species. If the protein evolved at the same rate in the two lineages, the number of amino acid changes between the common ancestor (O) and A should equal the number of changes between O and B (i.e. $d_{AC} = d_{BC}$). If not, then a global molecular clock, assuming rate constancy, cannot be applied to date the age of the common ancestor of A and B.

In practice, there are three possible difficulties with the exclusion approach. First, the RRT is sensitive to the choice of the in- and outgroups: the ingroup taxa A and B must be closest relatives and outgroup C must not be too far removed. Second, some RRT variants have a low power for the kind of data typically used, and so only dramatic departures from rate constancy are likely to be detected (Scherer 1989; Tajima 1993; Rambaut & Bromham 1998; Robinson et al. 1998; Bromham et al. 2000). This low detection power has serious implications, because any rate variation that remains undetected can result in consistently biased date estimates. Third, the exclusion approach is practical only if rate variation is the exception rather than the rule; otherwise, a large proportion of the sequences in a dataset has to be excluded.

Until the beginning of the 21st century, most rate estimates were based on pairwise distance methods with simple fixed calibration points (Wray et al. 1996; Kumar & Hedges 1998; Nei et al. 2001). In response to the difficulties just mentioned, methods assuming variable rates have been developed since then, including local clocks (Yoder & Yang 2000), nonparametric approaches (Sanderson 1997, 2002), and Bayesian parametric models (Thorne et al. 1998; Huelsenbeck et al. 2000; Kishino et al. 2001; Aris-Brosou & Yang 2002, 2003; Drummond et al. 2006). Among these methods, the Bayesian parametric approach offers the opportunity of exploring a wide diversity of alternative models, each of which corresponds to specific assumptions concerning the shape of the tree, the way the rate of substitution changes with time, and the way calibration nodes are constrained.

Local molecular clock methods: methods using a small number of rates

These methods assume that different parts of the phylogeny are characterized by different rates or 'local molecular clocks'. As long as the number of different rates assigned is small, they can be jointly estimated with the divergence times, as it is done

with a single fixed rate. If the number of rates is too large, however, then the rates and dates become 'non-identifiable' (i.e. an infinite number of rate and date combinations are equally probable).

The quartet dating method

The 'quartet method' (Rambaut & Bromham 1998; Bromham et al. 1998) is one of the simplest local clock methods. This method, based on that presented by Cooper & Penny (1997) but performed within a maximum-likelihood framework, combines two pairs of species, each of which has a known date of divergence. A rate can be estimated for each pair, and this enables the date of the divergence between the pairs to be estimated (Fig.4a). The rate change is arbitrarily placed at the midpoint root; therefore rate tests are used to exclude quartets in which members of the pairs have significantly different rates. Although the quartet method neatly avoids problems of topological uncertainty (because groups with undisputed relationships can be chosen), it is difficult to combine estimates from multiple quartets in a meaningful way, unless they are phylogenetically independent (Bromham et al. 1998). Furthermore, the method does not avoid the difficulties associated with the other rate-testing approaches mentioned above (i.e. the low power of tests, and the necessity of excluding large amounts of data). A program called QDATE had been written to perform these analyses but is not available anymore.

Local molecular clocks method

A related approach that avoids these difficulties is the local molecular clock method of Yoder & Yang (2000), which builds on the work of Kishino & Hasegawa (1990). In this method, a few rate classes are assigned to portions of the whole rooted tree (Fig.4b). The motivation here is that the clock hypothesis is likely to hold for closely related species. The assignment of these rate groups relies on the effective identification of anomalous lineages or groups, for example, by rate testing (e.g. via RRT) or the use of external knowledge (Yoder & Yang 2000; Kishino & Hasegawa 1990). Alternatively, rates can be assigned after informal preliminary examination of branch length estimates obtained without assuming rate constancy. However, the fact that the placement of the different local molecular clocks on the tree is left to the user's discretion can affect the inference (Aris-Brossou 2007). Another drawback of this approach, as Yoder & Yang (2000) point out, is that using the data to assign rate

placements should preclude the use of the same data to formally test the adequacy of those placements, which however is done in practice. In 2003, Yoder & Yang extended their previous likelihood models of local molecular clocks to accommodate multiple calibration points and multiple genes. The new models therefore account for heterogeneity among different genes in evolutionary rate and in substitution process; they are implemented in the PAML (Yang 1997) package (http://abacus.gene.ucl.ac.uk/software.html). It is possible to calculate local molecular clocks with R8S (Sanderson 2003) and BEAST (Drummond et al. 2007), although this is not the main purpose of these software.

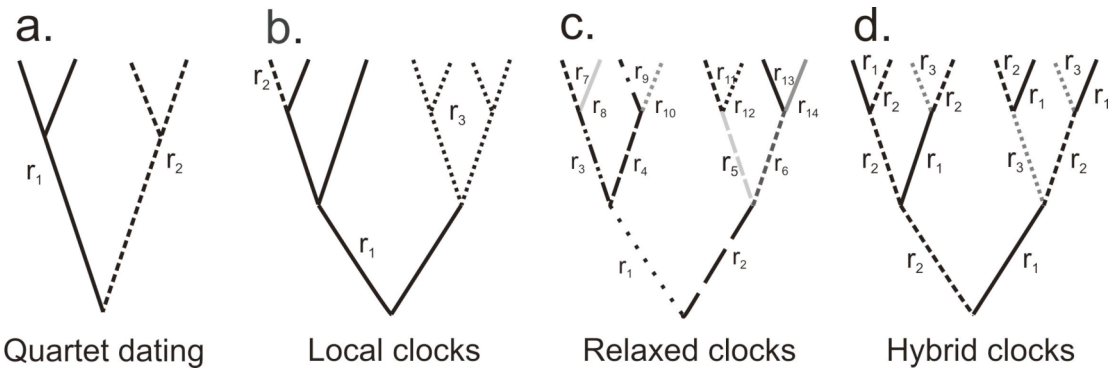

Figure 4: **Examples of variable-rate molecular-dating methods.** Each of the trees (a) to (d) shows an example phylogeny to be dated, and indicates where different rates, denoted as r_i, have been assigned to different branches. In the quartet method (a), a tree of four species is used, and both internal nodes must have external calibrations in order to date the basal node. In the local molecular clock method (b), a complete phylogeny is used, and placement of the additional rates, here r_2 and r_3, relies on the effective identification of 'anomalous' lineages. Methods allowing many different rates (c) must specify prior expectations for the value of each rate, r_i. According to the statistical framework used, these expectations are expressed either as a penalty function, or as a Bayesian prior. In Yang's combined method (d), the placement of a small number of rate classes (three in the example shown) is determined with the aid of prior expectations, as in (c). Each rate is then estimated as in (b). Modified from Welch & Bromham (2005).

Methods using many rates: I. Nonparametric approaches

Whereas local molecular clock methods rely on rate changes being relatively infrequent, other methods have been developed in which the rate can change many times (Fig.4c). To avoid problems of non-identifiability, these methods must rely on strong a priori assumptions about the way in which rates change over the tree. Sets of rates that conform to these assumptions are then favored during the estimation procedure. Within this broad class of methods, several approaches can be identified.

The first approach, wholly nonparametric, was introduced by Sanderson (1997) as "nonparametric rate smoothing" (NPRS). In this method the prior assumption about rates is that small changes in rate are more likely than are large changes. This assumption is embodied in a penalty function ($P = (r-r_p)^2$) minimized during the estimation (least square-smoothing criterion) in such a way that the model applies the least penalty when the rate of the branch under consideration (r) equals the rate of its parental branch (r_p). This method, however, tends to systematically apply rapid rate fluctuations on short branches of a tree which does not always reflect reality.

In a later implementation (Sanderson 2002), the rate-smoothing function used in NPRS (P) is combined with a likelihood model of branch lengths (L), thus generating a 'penalized likelihood' estimate. It permits specification of the relative contribution of the rate smoothing (P) and the data-fitting parts (L) of the estimation procedure using a smoothing parameter (λ). The user can set any level of smoothing (λ) from severe (high penalty for rate changes), which essentially leads to a constant molecular clock, to effectively unconstrained (low penalty for rate changes). The key to this approach is to find an objective method for selecting an optimal level of smoothing, therefore cross-validation procedures are provided for finding the optimal level of smoothing (Near & Sanderson 2004). In addition, the dates of some nodes can be fixed or constrained to lie within a given range (usually determined by fossil evidence). Penalized likelihood always outperforms NPRS, which tends to allow too much rate variation and thereby loose predictive power.

Methods using many rates: II. Hybrid methods between local clocks and rate smoothing

In many ways, local clocks and rate-smoothing methods are complementary: the first one deals with rare and large changes in rate, the other with small and frequent changes. There are two difficulties with local clock models. First, the number of local clocks has to be chosen, and second, local molecular clocks must be placed on a given phylogeny. However, in contrast to rate-smoothing methods, these models are expected to be able to accommodate high levels of rate heterogeneity. The method of Yang (2004) and its later extension (Aris-Brosou 2007) can be viewed as hybrid

approaches, combining elements of both rate smoothing and local molecular clocks, and present solutions to avoid the subjective choice of placing the local clock on the phylogenies.

Yang's heuristic rate-smoothing method (2004) consists of two quite separate stages. The first stage is 'ad hoc rate smoothing' (AHRS). This stage yields a distinct rate estimate for each branch of the tree. Using these estimates, branches with similar rates are placed in the same group, creating a small number of rate groups. The second stage is a conventional local molecular clock estimation (Yoder & Yang 2000) using the rate group placements determined by the first stage (Fig.4d).

Aris-Brosou (2007) recently presented four new ad hoc methods that improve on Yang's AHRS algorithm (2004). Rather than the automatic placement of local molecular clocks in the AHRS algorithm, these methods provide researchers with a means to determine how many clocks should be used to analyze their data. Even though these new maximum likelihood hybrid methods seem to perform better than penalized likelihood, and almost as well as uncorrelated Bayesian models (see below), they still tend to underestimate the actual amount of rate change. Moreover, these hybrid methods appear to present potential limitations (Aris-Brosou 2007), among which the difficulty to obtain confidence intervals, because the uncertainties about model parameters and calibration points are disregarded. The AHLC software can be found at http://www.bioinformatics.uottawa.ca/stephane/.

Methods using many rates: III. Bayesian parametric models

This approach, pioneered by Thorne et al. (1998) (Fig.4c), uses Bayesian statistics, a framework in which prior beliefs about parameters, reflecting the range of plausible evolutionary histories, are exploited in the estimations (e.g. Holder & Lewis 2003). The prior assumptions about parameters are expressed in a formal probability distribution (e.g. normal, log-normal, gamma, exponential,…) so patterns that depart from these assumptions are assigned lower probability values. Posterior distributions (the resulting values) are compromises between prior distributions and data information. The fact that Bayesian statistics requires that prior probabilities are specified for all divergence dates and rates along branches can be problematic if the information carried by the priors is erroneous. In some studies, the 'date priors' seem to have been a major determinant of the date estimates obtained, even though their

distribution cannot be adequately justified in terms of the prior knowledge available (Kishino et al. 2001; Douzery et al. 2004; Pérez-Losada et al. 2004; Yoder & Yang 2004; Welch et al. 2005). The influence of the prior diminishes, however, with the amount of information carried by the data used (Huelsenbeck et al. 2002; Holder & Lewis 2003; Douzery et al. 2004). In all these methods, different combinations of all parameter values are sampled by using a Markov Chain Monte Carlo (MCMC) method using a scheme that is more likely to choose parameter values that explain the data well. MCMCs are sampled until stability is reached, i.e. the optimal parameter values are found. The MCMC exploration of parameter space enables the algorithm to estimate a posterior distribution of the parameters: the dates, the rates of substitution (depending on the method), the tree shape (depending on the method), and so on.

Many alternative probabilistic models of clock relaxation have been proposed in the Bayesian framework. They differ in various respects: some describe the rate as a continuous process with the priors for rate modeled as a geometric Brownian motion (e.g. Kishino et al. 2001; Rannala & Yang 2007), others as a compound Poisson process (Huelsenbeck et al. 2000), and still others do not explicitly describe the rate process itself, but rather directly assign a mean rate parameter to each branch (Aris-Brosou and Yang 2003; Drummond et al. 2006). Models also differ in their dynamics: rates may be autocorrelated (Kishino et al. 2001; Aris-Brosou and Yang 2003, Rannala & Yang 2007) or not (Drummond et al. 2006; Rannala & Yang 2007). Finally, different priors have been proposed on the set of divergence times, including a Yule process (Thorne et al. 1998; Drummond et al. 2006), a generalized Dirichlet distribution (Kishino et al. 2001; Thorne & Kishino 2002) and a birth-and-death process (Yang & Rannala 2006; Rannala & Yang 2007; Inoue et al. Submitted).

Hulsenbeck-Larget-Swofford model for rate evolution

In the method developed by Huelsenbeck et al. (2000), rate change events are modeled as step-like processes on a phylogeny: the changes can occur anywhere on the tree, not just at a set of predetermined points (such as at the internal nodes as in other methods [*e.g.* Sanderson 1997; Thorne et al. 1998]), which should render the method less sensitive to taxon sampling. It works as follows: positions on the tree at which substitution rate changes occur according to a Poisson process with parameter λ; when a rate change occurs, the rate of substitution just prior to this event (r_p) is multiplied by a gamma-distributed random variable to produce a new substitution rate

after the event (r).

The compound Poisson process can be regarded as a biologically meaningful way to describe rate change. For example, any process that rapidly changes the substitution rate (such as a change in the proofreading mechanisms, a change in selective constraints, or a rapid change in the generation time) can be approximately described using a step function. Conversely, the gamma distribution used to modify ancestral rates is not so easily justified using as an argument biological plausibility. Other methods for modifying rates may work just as well (Huelsenbeck et al. 2000).

The Bayesian method of Huelsenbeck et al. (2000) contains a model of rate change that, like most similar methods, penalizes large changes. Uniquely, however, the model enables the number of rate changes to vary during the estimation. Because there are two distinct kinds of constraint on the rates, with penalties applied to the number of rate changes and to their magnitude, one or the other can be given greater prominence. To test the sensitivity of the date estimates to the different weightings, Huelsenbeck et al. (2000) applied their method while making different prior assumptions about the number of rate changes. They showed that the same data could be interpreted as containing many changes of small effect (the assumption governing the rate-smoothing approaches) or as containing a few changes of large effect (the assumption behind the local molecular clocks approaches); the date estimates for most nodes varied little in the two cases. However, Rannala (2002) showed that the model used by Huelsenbeck et al. (2000) was overparameterized, and no follow up study has been published until now.

Aris-Brosou & Yang's method

Because several models of rate change have been proposed, Aris-Brosou & Yang (2002, 2003) decided to extend and evaluate them, i.e., the lognormal model (Thorne 1998) and its variant (stationary lognormal, Kishino 2001), along with the gamma, the exponential, and the Ornstein–Uhlenbeck processes (Aris-Brosou & Yang 2002). They found that the rate and date estimates varied little from model to model, although the posterior Bayes factor indicated that the Ornstein–Uhlenbeck process outperformed the other models. The major differences could only be seen between local clock versus relaxed clock models. This new model is implemented along with four others in a program called PHYBAYES (http://www.bioinformatics.uottawa.ca/-stephane/). It presents the advantage of using a new, and maybe better, model of rate

changes and sequence data to directly estimate branch lengths. However, this program also has some weaknesses, e.g. the use of a single fixed calibration point.

Thorne & Kishino's method: branch autocorrelation and multiple fossil calibrations with "hard" bounds

Thorne et al. (1998) relaxed the molecular clock by assuming that rates change across speciation events in an autocorrelated manner. This is based on the idea that the factors responsible for divergence of rates among lineages (e.g. population size, generation time, fidelity of DNA replication) may themselves be quite similar among closely related lineages. This method assigned rates to descendant lineages by sampling rates from a normal distribution with a mean being the logarithm of the rate of the ancestral branch, and with a variance equal to the time difference between the midpoint of the ancestral and descendant branches (Δt) multiplied by an autocorrelation parameter (ν). Consequently, when branches are long (Δt) and ν is big, the autocorrelation between branches will be low. In other words, when $\nu = 0$ the global clock is applied. In this case the time priors are modeled by a flexible generalized Dirichlet distribution, cutting the path from the root to the tips in various fragment length. With this method it is possible to add some time constraints. Each constraint is defined by a uniform distribution marked off by a lower and/or an upper bound. This is called "hard bound" because any date outside the distribution has a null probability. This is one of the major differences between this method and the following ones.

The method is implemented in the computer program MULTIDIVTIME (Thorne et al. 1998; Kishino et al. 2001; Thorne & Kishino 2002) that can be uploaded at http://statgen.ncsu.edu/thorne/multidivtime.html

Drumond & Rambaut's method: uncorrelated branches and multiple fossil calibrations with "soft bounds"

Although prior distributions and implementation details vary, the Bayesian approach and penalized likelihood both have in common to smooth or minimize rate variation over evolutionary time by means of an autocorrelated process. However, according to Drummond et al. (2006), the autocorrelation of rates is not justified whenever the largest component of rate variation is not due to inherited factors. This

would be the case as one looks over smaller and smaller timescales; the differences in inherited factors become smaller relative to the variance caused by stochastic and non-heritable factors (such as environmental or chance events). At the other extreme, over very long timescales, we might expect so much variation in the inherited determinants of rate that the autocorrelation from lineage to lineage begins to break down, especially with sparse taxon sampling. However, it is difficult to predict where the demarcations between these effects are, and thus to specify what the degree of autocorrelation will be. In response to these arguments they introduced a new approach incorporating models of uncorrelated rate change (Drummond et al. 2006; Drummond & Rambaut 2007).

In contrast to previous methods, this Bayesian method estimates the divergence times, the topology of the tree and the rates, all as part of the same calculation. Even though they are not correlated, rates are not completely unconstrained; they are assumed to be drawn independently from an identical discretized distribution (log-normal or exponential). The continuous distribution of branch rates is approximated using as many discrete rates as there are branches on the tree. Priors about the divergence times are modeled following a Yule process. This method was one the first to allow for uncertainty in the dates attributed to calibration points by modeling calibration informations using different probability distributions (exponential, normal and log-normal).

According to the authors, this method, implemented in the computer program BEAST (Drummond & Rambaut 2007, http://beast.bio.ed.ac.uk/), is phylogenetically more accurate and precise than the traditional unrooted method (PAUP*, MrBAYES), while adding the ability to infer a timescale to evolution. However, consensus has not be reached yet as Lepage et al. (2007) found a better fit of the autocorrelated model, while Rannala and Yang (2007) pointed out that the uncorrelated method tends to underestimate possible positive autocorrelations in rates across branches because the same rate cannot be assigned to two connected branches. So they conclude that Drumond & Rambaut's method, technically, does not appear to be a correct implementation of the variable-rates models.

Yang & Rannala's method: correlated/uncorrelated branches and multiple fossil calibrations with "soft bounds"

Yang & Rannala (2006) highlighted the critical importance of fossil calibrations to molecular dating and the need for probabilistic modeling of the knowledge contained in the fossil record (fossil depositions, preservations and sampling) to provide statistical summaries of information concerning species divergence times. Up to then, only Drummond et al. (2006) had also tried to take into account the uncertainty associated with paleontological dates. Yang & Rannala (2006) first developed a "soft bound" approach: in this approach, calibration ages are not uniformly distributed between two "hard" bounds — the probability that the calculated date falls outside the interval being zero — but more flexible distributions are implemented to allow the calculated age (even though with a small probability) to fall outside the defined interval.

A birth-death process specifies priors for divergence times. In their first study, the impact of "soft" vs. "hard" bounds was assessed through a strict molecular clock (Yang & Rannala 2006) that got relaxed in a second step (Yang & Rannala 2007). In a very recent study (Inoue et al. submitted) the statistical distribution describing the prior knowledge about the fossil record is refined and can take into account the quality of the calibration (poor or accurate). This method has been implemented in the MCMCTREE v4.2 software distributed with the PAML package (http://abacus.gene.ucl.ac.uk/software.html).

Methods using many rates: IV. Cutler's method

A quite distinct approach to incorporating rate variation was developed by Cutler (Cutler 2000). As this author pointed out, the observation that sister branches can have widely variable numbers of substitutions could have two distinct explanations: (i) that different rates of molecular evolution characterize the lineages; or (ii) that the process of molecular evolution is identical in both lineages, but simply has a high variance (owing, perhaps, to substitutions tending to cluster in time); if this were the case, random sampling alone might have produced the observed variation. Furthermore, possibilities (i) and (ii) might be fundamentally indistinguishable in some cases. Unlike all methods discussed so far, which are based on assumption (i), Cutler's method accepts assumption (ii), so that all lineages are assigned the same

basic evolutionary rate, but the process can be highly variable such that rapid bursts of substitutions might occur on some lineages.

Cutler's method resembles rate smoothing in that departures from rate constancy are in effect penalized during the estimation. However, unlike some other models, Cutler's method does not assume that bursts of substitutions are most likely to occur on closely related lineages or, alternatively, that rapidly evolving lineages are most likely to give rise to other rapidly evolving lineages. Rather, the method penalizes departures from the overall mean rate of the tree, regardless of the smoothness with which the changes take place.

Comparing the methods

A number of empirical and simulation studies have compared the various methods as a whole or have focused on specific parameters in order to estimate their influence on divergence time calculations. In particular, it is not totally clear whether autocorrelation is a feature that relaxation models should always take into account (Drummond et al. 2006). A few controversies have also arisen about possible biases in certain cases due to the relaxation model itself or to the prior on divergence times (Blair and Hedges 2005; Welch et al. 2005). More generally, the relaxation model seems to have a non-negligible influence on the resulting divergence dates (Perez-Lozada et al. 2004; Smith et al. 2006), a fact that does not help sorting out the still open controversies between fossil and molecular datings (e.g. Bromham et al. 1999; Smith and Peterson 2002). As these controversies sometimes bear on the very choice between recent and explosive radiation versus ancient and progressive diversification scenarios, for example in the case of metazoans (Smith and Peterson 2002; Aris-Brosou and Yang 2003; Douzery et al. 2004; Peterson et al. 2004; Blair and Hedges 2005; Welch et al. 2005) and mammals (Bromham et al. 1999; Springer et al. 2003), deciding between alternative models of clock relaxation is obviously urgent as there seems to be some arbitrariness in the choice of the model of rate evolution.

As already mention earlier, Aris-Brosou & Yang (2002) implemented and compared different models of autocorrelated rate change over time in a likelihood framework, focusing on two points: the effect of the model of rate change and the effect of the parameterization of each model to relax the clock. The rate and date estimates varied little from model to model. However, the Ornstein–Uhlenbeck

process (OUP) seemed to perform better than the other models. In a subsequent article, the same authors (Aris-Brosou and Yang 2003) used two of these models (the exponential and the OUP) to assess molecular estimates of basal metazoan divergences. Their results were in accordance with the dates extracted from the fossil record and younger than previous molecular estimates. However, Welch et al. (2005) heavily criticized these results and showed that they were strongly influenced by questionable a priori assumptions. In principle, there is nothing wrong with assigning informative priors, if there is good reasons to believe the assumptions that they embody, or if there is enough data to override the false assumptions. Welch et al. (2005) suggested, however, that Aris-Brosou and Yang's priors did not accurately reflect the biological situation being modeled, and was not "overridden" by the data.

Lepage et al. (2007) tried as well to understand which models among the various available ones yield reliable dates. Concerning the relaxation process, in contrast to Drummond et al. (2006), their comparisons give a nearly unanimous answer in favor of autocorrelated models. They propose that this kind of model should be used on a systematic basis. On the other hand, concerning the choice of the prior on divergence times, no clear guidelines can be deduced from their comparisons as the results turn out to depend strongly on the data sets.

Explaining the different outcomes of dating methods is complicated, because they differ not only in the way in which they deal with rate variation, but also in their treatment of the other aspects of molecular dating, such as branch length estimation, the use of external fossil constraints and the statistics involved. A few of these differences are necessary corollaries of the assumptions made about evolutionary rates, but most are more or less arbitrarily associated with different implementations. There has been discussion about the validity of the different models used to estimate divergence times from molecular data, but few empirical tests have been reported in taxonomical groups with a good fossil record. Smith et al. (2006) is one of them; in their study on echinoderms they have compared the fully parametric Bayesian method of Thorne et al. (1998), the nonparametric rate smoothing (NPRS), and the semiparametric penalized likelihood (PL) methods of Sanderson (2002), and also included a dating method imposing a strict molecular clock, the Langley–Fitch method (LF; Langley & Fitch 1974), to evaluate its performance for comparison. Molecular estimates of divergence times derived from applying both molecular clock and relaxed molecular clock models are concordant with estimates based on the fossil

record in up to 70% of cases, with most concordant results obtained using Sanderson's semiparametric PL method and a logarithmic-penalty function. However, the quality of the results depended greatly on the number of fossil calibrations used in the analyses: when only the basal node of the ingroup was fixed, all methods performed poorly, whereas with the addition of four local calibration points, all molecular estimation methods performed reasonably well. They explain the better results of the Sanderson's method (2004) by the fact that the method of Thorne et al. (1998) can only use simple models (e.g. F84), but it builds these models into the likelihood calculations, whereas Sanderson's (2004) software (R8S) relies on trees with branch lengths that must be previously estimated using other software (e.g. PAUP*, MrBAYES). Sanderson's method (2004), in this way, allows using more complex models and hence more accurate branch length estimations. If this is true, then we should expect that the Bayesian method of Thorne et al. (1998) would improve as more models are integrated into it.

On the other hand, Perez-Losada et al. (2004) estimated dates that varied considerably within and between approaches depending on the calibration points. Highly parameterized local clock models that assumed independent rates for confamilial or congeneric species (Drummond et al. 2006) generated the most congruent estimates among calibrations, and agreed more closely with the fossil record. Reasonable estimates were also obtained under the Bayesian procedure of Kishino et al. (2001), but using multiple calibrations. Most of the dates estimated under the Bayesian procedure of Aris-Brosou & Yang (2002) and the penalized likelihood method using single and/or multiple calibrations were inconsistent among calibrations and did not fit the fossil record.

In conclusion, when using molecular data to estimate divergence times, modern parametric and semiparametric approaches assuming rate heterogeneity generate realistic divergence time estimates in the great majority of cases as long as a realistic model of rate variation is applied (see Welch et al. 2005). However, according to Inoue et al. (submitted) the major parameter influencing the posterior time estimation is the prior on times, which incorporates fossil-calibration information, therefore we will focus on this parameter in the next section.

UNCERTAINTIES OVER THE DATES OF CALIBRATION POINTS WILL BIAS AGE ESTIMATES

Each branch length on a molecular phylogenetic tree is the product of two components: the substitution rate and the time elapsed. With molecular data alone, we are unable to separate the contributions of these two components; therefore, in order to establish an absolute (geological) time scale, it is necessary to introduce some form of additional calibrating information. This can take several forms. For instance, this can be done by importing a known substitution rate, estimated in an independent study for the taxon in question (Fig.5a), or by including in the analysis organisms with known ages sampled at different sampling time (heterochronous sequences, e.g. from subfossils, museum specimens or viruses) (Fig. 5b). In the first case this can only be done at a really low phylogenetic level, and the information needed in the second case is relatively rare. Classically, however, the fossil record or specific geological events provide the time scale for evolutionary history. Evolutionary time scales based on molecular clocks have been controversial, especially in their early development, because they often clashed with time estimates from the fossil record. In some cases, molecular date estimates were up to twice as old as paleontological dates. However, although it is true that molecular dates are often too old, due to statistical bias, paleontological dates are often too young, due to missing fossils (Benton & Ayala 2003). Methodological problems linked to modeling molecular evolution have been exposed in the previous section; the present section deals with the paleontological uncertainties.

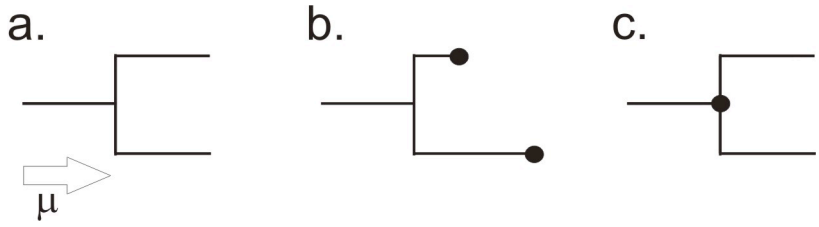

Figure 5: Different methods for calibrating estimates of substitution rates and divergence times. (a) Importation of an independently derived rate, μ. (b) Inclusion of sequences with distinct, known ages. (c) Point calibration: fixing the age of a node to a specific value based on independent evidence. Modified from Ho (2007).

Paleontological uncertainties

In order to assess the evolutionary timescale, the molecular clock needs to be calibrated using external information about the geological ages of one or more nodes in the phylogeny to convert the estimated branch lengths into geological times. This information is typically based on the fossil record, but the formation of landmasses (e.g. islands) or separation of continents can as well be used as calibration. Calibration points – from geological or paleontological information -have often been used without considering the errors associated with them, which has recently attracted strong criticism (Graur & Martin 2004; Heads 2005). Indeed, ignoring the uncertainty in the date attributed to a calibration point at the start of the analysis leads to date estimates with overly optimistic small confidence intervals (van Tuinen & Hadly 2004a, b). However, assessing the uncertainty is not easy, as it can be due to several parameters (Benton & Donoghue 2006; Donoghue & Benton 2007):

- Phylogenetic topology: the phylogeny in which calibration points are assigned should be correct and display robust support values for the nodes to be calibrated, as well as for the nodes above and below.
- Identification: all the fossils used in the analyses should be correctly identified and assigned to their lineages.
- Fossil record sampling: the oldest known fossil will not be the earliest member of a lineage, and the oldest actual fossil is unlikely to ever be sampled. Identifying the oldest actual fossil representative of a clade is difficult because: (i) it emerges at a single point in time and space; (ii) the earliest representative of a clade will invariably lack fossilizable apomorphies because of its recent emergence; and (iii) fossils are usually incomplete, and so it can be difficult to determine whether the absence of clade-specific diagnostic characters reflects the nature of the organism or of its fossilization history.
- Exact age-date assignment (absolute dating): good radiometric dates should be assigned to the fossiliferous horizons, and their errors should be estimated.
- Correlation (relative dating): dating of the fossils used in calibration is rarely direct; dates are usually assigned through correlation of the rock section from which the fossil was recovered to an other section for which absolute age dates are available. This process is not always simple and can be prone to error.

Donoghue & Benton (2007) conclude that fossils cannot provide accurate estimates of evolutionary splitting events, but the oldest fossil(s) of a crown group can provide a firm minimum age constraint on such events (Fig.6).

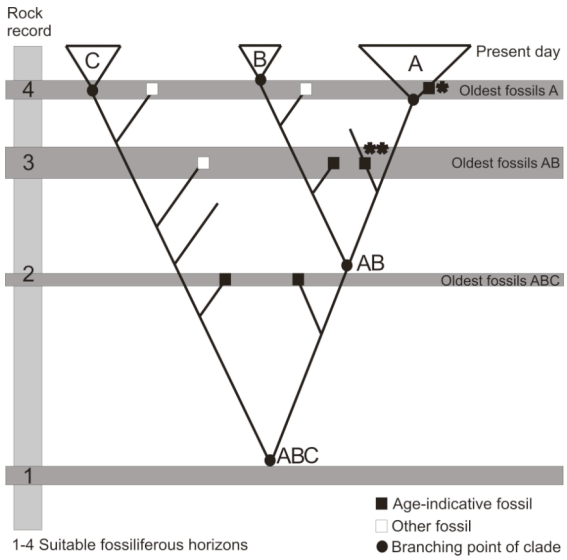

Figure 6: Definitions of terms in assigning fossils to clades. A crown clade consists of all living species in a clade and their most recent common ancestor (triangles A, B and C); this is preceded by a stem lineage of purely fossil forms that are closer to that crown clade than to another crown clade. The divergence or splitting point between a species in clade A and a species in clade B is the point AB. This is older than the points of origin of crown clades A and B. Fossils may belong to a crown clade (*) or to a stem lineage (**), and cladistic evidence should indicate which is the case. In this figure, four fossiliferous horizons are the source of all relevant fossils. Minimum constraints are indicated by the oldest fossils for ABC, AB, and A (black squares). Fossiliferous horizon 1 that contains no fossils assignable to the clade ABC marks a maximum constraint (soft bound) on the age of the clade. Fossiliferous horizon 2 marks a maximum constraint on the age of clade AB. Modified from Benton & Donoghue (2007).

Modeling the paleontological uncertainties

Early applications of the molecular clock to date species divergences typically use a single calibration point, treated as known without error (Graur & Martin 2004; Hedges & Kumar 2004) (Fig.5c). In theory, it should be possible to overcome the inherent uncertainty of taking a single calibration date by using several different calibration points, but this strategy has its own associated problems. Calibration points are essentially used to infer the rate of the molecular clock, yet we know that this rate varies across the tree. The use of multiple calibration points will, therefore, only be effective and reliable if we specify correctly how the rate varies across the tree, a topic still in progress. We can be relatively confident in the dating of clades that have internal calibration points, especially those with calibration points directly

above the node (branch point) of interest and whose ages reached a consensus in the paleontological community. Unfortunately, it is rare to have such closely spaced and relevant calibration points. More often, attempts are made to extrapolate from rates estimated from one part of a phylogenetic tree to another, or from one time period to another. In this case, the information from multiple calibration points might indicate that the rates vary, so that a properly calculated extrapolation would only produce imprecise estimates (Soltis et al. 2002; Smith et al. 2006).

One approach is to model diversification pattern and extinction/preservation probability over time (Foote et al. 1999; Tavaré et al. 2002). However, even though modeling approaches are particularly well suited to testing molecular clock estimates, they are perhaps too assumption-laden to use in molecular clock calibration (Donoghue & Benton 2007).

Another approach to account for paleontological uncertainties is phylogenetic bracketing, which includes not only minimum, but also maximum constraints on the timing of a branching event using the date of the preceding and subsequent branching episodes (Reisz & Muller 2004; Muller & Reisz 2005). Broader constraints can be derived using the earliest stem-member of the overall clade to provide a maximum constraint, and the earliest member of the crown group to provide a minimum constraint (van Tuinen & Hedges 2001; van Tuinen & Hadly 2004a); propagated errors can then be placed on both these dates to provide the overall extent of the bounds.

Another approach to account for paleontological uncertainties is phylogenetic bracketing, which includes not only minimum, but also maximum constraints on the timing of a branching event using the date of the preceding and subsequent branching episodes (Reisz & Muller 2004; Muller & Reisz 2005). Broader constraints can be derived using the earliest stem-member of the overall clade to provide a maximum constraint, and the earliest member of the crown group to provide a minimum constraint (van Tuinen & Hedges 2001; van Tuinen & Hadly 2004a); propagated errors can then be placed on both these dates to provide the overall extent of the bounds.

Some studies have attempted to use statistical distribution to describe uncertainties in the fossil dates. One of the first assumed that the calibration age was uniformly distributed between two bounds—the probability that the date falls outside the interval is then zero (Thorne et al. 1998) and an equal probability is assigned to all

values within the interval. (Fig.7a). But this ‘‘hard’’ bound, such as that imposed by a uniform prior, often overestimates the confidence in the fossil record. Indeed, fossils can only underestimate the actual date of a particular evolutionary event because the oldest fossil is necessarily younger than the origin of its group, which means that they often provide good lower bounds (i.e., minimum node ages), but not good upper bounds (maximum node ages). Consequently, one may choose to use an unrealistically high upper bound to avoid precluding an unlikely (but not impossible) ancient age for the clade. However, this strategy is problematic as the bounds imposed in the prior may influence the posterior time estimation in an MCMC framework.

For these reasons, maximum constraints should be estimated using more flexible distributions and “soft” bounds. Drummond et al. (2006) and Yang & Rannala (2006) have implemented the possibility to treat calibration times with parametric prior distributions. In BEAST (Drummond & Rambaut 2007) these distributions can be normal (models a non-directional uncertainty), lognormal (one can model that the actual divergence event is most likely to have occurred some time prior to the earliest appearance of fossil evidence) or exponential (the probability decreases with a growing discrepancy between estimated nodal age and the age of the calibrating fossil) (Fig.7b,c,d). They provide a more realistic assessment of the uncertainty associated with the fossil record. In the framework they present, the tree itself is being sampled and thus the age of a particular internal node cannot be defined. Instead they specify the age, or the prior distribution of age, for the most recent common ancestor of a set of taxa. Every time a new tree is proposed in the MCMC chain, the most recent common ancestor of the specified taxa is located in the tree, and the prior probability of the age of this node is used to assess the acceptance probability of the proposed tree. Yang & Rannala (2006) and Inoue et al. (submitted) have also implemented a Bayesian MCMC algorithm in which fossil dates are set as distribution: uniform distribution with soft bounds, gamma distribution (allows roughly to model the priors as exponential and log-normal distributions) and minimum bound density (Fig.7e,f,g). In this last distribution, knowledge about the accuracy of the fossils used in the analyses can be incorporated (through the parameters c and p). In all these cases, the shape of the various distributions should be based on a careful assessment of the fossil and geological data on which the bound is based.

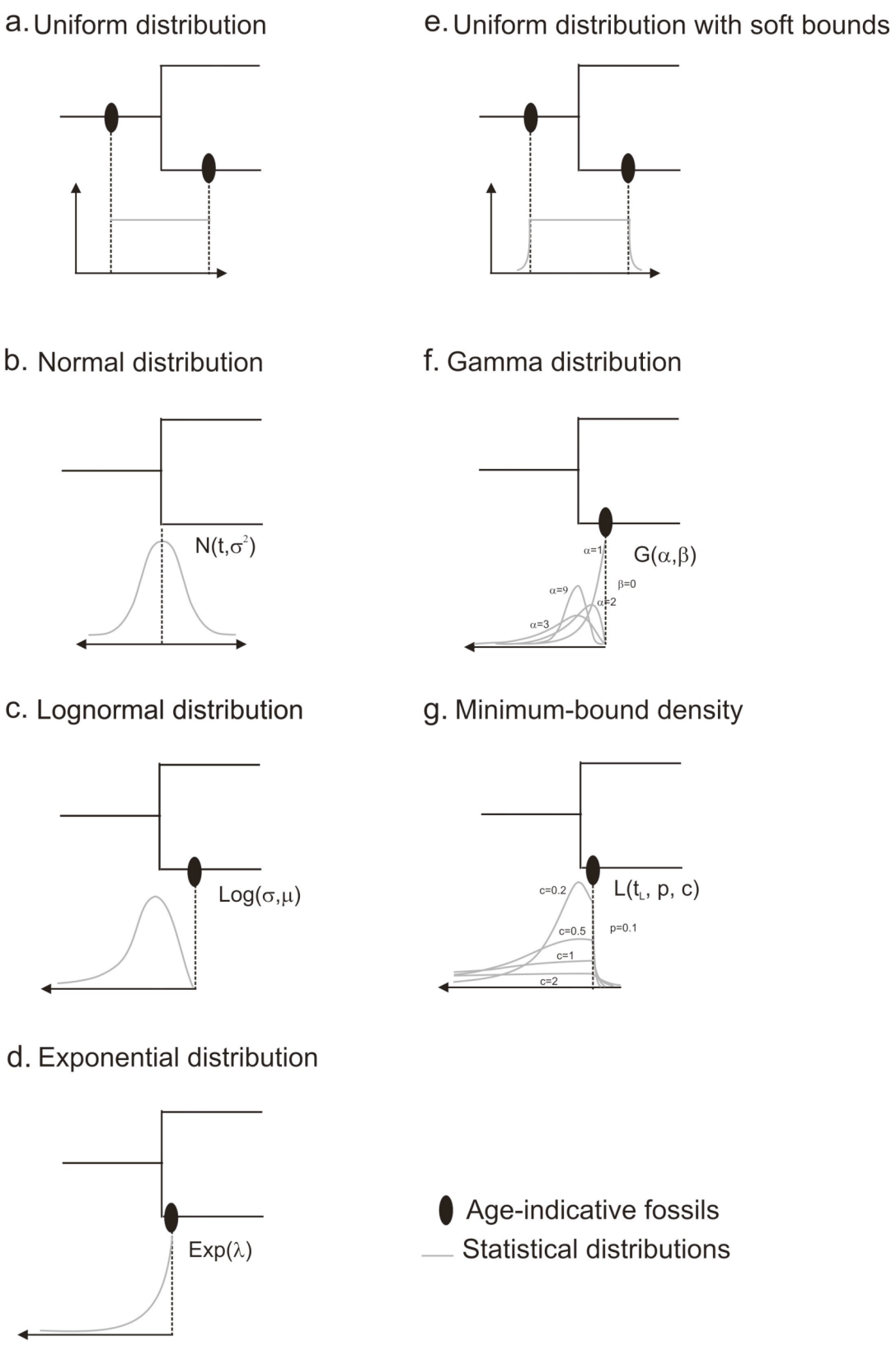

Figure 7: Parametric distributions of the age of a constrained node. (a) Uniform distribution (Kishino et al. 2001). (b-d) Distributions implemented by Drummond et al. (2006): (b) normal distribution, (c) lognormal distribution, with a rigid minimum bound, (d) exponential distribution, with a rigid minimum bound. (e-g) Distributions implemented by Yang & Rannala (2006) and Inoue et al. (submitted): (e) uniform distribution with soft bounds, (f) gamma distribution, with a rigid minimum bound, (g) minimum-bound density.

In conclusion, when fossils are consistent with each other and with the molecular data, and the posterior time estimates are well within the prior bounds, soft and hard bounds produce similar results. However, when the fossils are in conflict with each other or with the molecules, soft and hard bounds behave very differently; soft bounds allow sequence data to correct poor calibrations, while poor hard bounds are impossible to overcome by any amount of data. In addition, soft bounds eliminate the need for ‘‘safe’’ but unrealistically high upper bounds, which may bias posterior time estimates, and they allow more reliable assessment of estimation errors, while hard bounds generate misleadingly high precisions when fossils and molecules are in conflict (Yang & Rannala 2006).

Testing congruencies between calibration points

Methods to test the congruencies between the fossil calibrations are evolving in parallel with the dating methods. They become more and more refined as the models of rate of evolution and the definitions of calibration constraints become more sophisticated. A few methodologies have been developed to test the consistency of dating results obtained when using different calibration points in a study. This way it is possible to keep the ones that presumably represent accurate age estimates, versus fossil calibration points that would be erroneous. However, we have to bear in mind that differences in results between calibration points can be due to strong rate heterogeneity among lineages that could not be accommodated by the presently available methods (Soltis et al. 2002).

One of the very first molecular studies assessing the reliability of fossil calibrations was conducted in mammalian phylogeny (Springer 1997). Adjusted Tamura-Nei distances were plotted versus paleostratigraphic divergence times, and this regression was used to assess the divergence times of the major mammalian orders. In this study, Springer already tried to correct for heterogeneity of evolutionary rate among lineages.

Until a few years ago, because of the generally small number of sequences available for a given gene, it was not always possible to use direct calibration points. Therefore, some authors used secondary calibration points, i.e. molecular dates calculated in studies for which a primary external calibration point was available, independently of the original external calibration point. These secondary calibrations

were usually derived from the widely used bird-mammal divergence event, dated then at 310 MYA. This way of doing has been highly criticized (Graur & Martin 2004) and a consistency test was designed for assessing the reliability of divergence date estimates based on such secondary calibration points (Shaul & Graur 2002). This test was based on the reciprocal assessment of one calibration point by the other using corrected distances between protein sequences.

By using multiple calibration points, one can potentially calculate confidence intervals around molecular age estimates (Smith & Peterson 2002). While this approach has its merit, it may also be the case that some fossil datings are so inaccurate that one would be better off eliminating them rather than including them in a multifossil calibration of a molecular phylogeny. In 2004, van Tuinen & Dyke, using a global clock, made a very simple cross-validation assessment by looking at the dating results given by each fossil calibration separately. Some of the calibrations gave consistently outlying time estimates. Later on, Near & Sanderson (2004), using a relaxed clock, have developed a test called 'fossil cross-validation'. This is a procedure used to identify the impact of different calibrations on overall time estimation. It can identify fossils that have an exceptionally large error effect, and may warrant further scrutiny. Practically, the calibration points are used, in turn, to calculate the age of all other calibration nodes and, depending on their ability to estimate molecular ages as closely as possible to the fossil ages, are ranked. The calibration point giving the most divergent molecular ages is removed first, and an F-test is calculated to test whether the removal of that fossil calibration results in a significant reduction in the variance of the log-transformed differences between molecular and fossil estimates of node age. The removal process goes on until the outcome of the F-test becomes not significant. Only the remaining fossil calibrations are kept in the analyses. For this test, fossil ages are considered either as fixed or as minimal ages. Another test developed by the same research group is the 'Fossil-based model cross validation' (Near & Sanderson 2004); it is an entirely different procedure that uses fossils to identify the optimal model of molecular evolution in the context of rate smoothing or other reference methods. In this thesis, we have followed the same rationale, but applied it in an opposite way. Instead of using one fossil calibration point to assess all the others, we have removed each calibration point in turn and re-estimated it using all the others (Poux et al. 2005, 2006, 2008). The ages of the fossils were defined as uniformly distributed within a window of time. When the estimated

molecular ages were not consistent with the paleontological ones, the calibration point was removed from the fossil calibration set.

Rutschmann et al. (2007) have extended the 'fossil cross-validation' method (Near & Sanderson 2004) to compare the internal consistencies among entire calibration sets formed by multiple fossils that can be attached to alternative calibration points. This way they can identify the fossil assignments producing the calibration sets internally the most consistent. This procedure allows addressing the question whether each fossil is assigned to the most reasonable node.

Building on the idea that fossil calibrations are neither fully correct nor incorrect, Sanders & Lee (2007) used Bayesian relaxed clock analyses with soft and hard calibration bounds to demonstrate that the bird-lizard and bird-crocodile divergences were relatively broadly spaced in time. The Bayesian approach they used recognizes that calibrations lie on a continuum between highly accurate (narrow bounds straddling the actual divergence) and very inaccurate (wide bounds that still do not encompass the real date). The concordant, reliable calibrations contribute most to the final date estimates (priors consistent with posteriors); while less reliable calibrations have less influence (priors inconsistent with posteriors). These last two methods differ from Near & Sanderson (2004) and Poux et al. (2005) in that there is no cut off to retain a subset of calibrations; they do not treat calibrations in an all-or-nothing fashion.

MOLECULAR DATING TO TEST BIOGEOGRAPHICAL SCENARIOS OF MAMMALIAN EVOLUTION

During this thesis work we mainly focused on the biogeography of four endemic mammalian groups that succeeded to colonize South America and Madagascar: primates, rodents, carnivores and tenrecs; (Fig.8). Biogeography is the study of the distribution of biodiversity over space and time. The patterns of species distribution can usually be explained by a combination of historical factors such as speciation, extinction, continental drift, glaciations, dispersal, etc... The classic descriptive biogeography born in the 19th century has been expanded by the development of molecular systematics, creating a new discipline known as phylogeography. This development allowed scientists to test theories about the origin

and dispersal of populations, such as island endemics. To reach this aim one needs reliable phylogenies and divergence times, a good fossil record, insights in the geographical distribution of the species involved, and knowledge of plate tectonic movements. All these informations can then be superimposed to reconstruct the biogeographical history of living organisms.

Phylogeny and divergence times of eutherian mammals

Mammals display a great diversity of form and function, but they share three characters not found in other animals: three middle ear bones, hairs and mammary glands. Until recently the systematics of mammals was exclusively the field of paleontologists and morphologists. The reconstruction of the phylogenetic relationships of mammals by means of morphological characters proved difficult because of their great morphological diversity and complex ecological adaptations. However, the current morphological consensus tree of the class Mammalia is pretty well resolved (Fig.8a) (Novacek 1992). It displays 26 orders, of which one comprises the Monotremata (egg-laying mammals: e.g. platypus), seven are grouped in the infraclass Marsupialia (pouched mammals: e.g. kangaroo), and 18 form the infraclass Eutheria (also called placentals). Resolving the mammalian tree by means of molecular data turned out not to be an easy task because of discrepant results and disagreements between researchers, depending on the genetic markers they favored (mitochondrial vs. nuclear DNA) (e.g. Killian et al. 2001; Janke et al. 2002; Reyes et al. 2000; Huchon et al. 2002; Springer et al. 2001). With only two exceptions (the artiodactyl and insectivore orders) the molecular data support the monophyly of each traditionally defined eutherian order (Fig.8b). The main differences between the morphological and molecular trees arise at the superordinal level, where molecules have remodeled the mammalian tree and strongly support the division of Eutheria into four groups: Afrotheria (e.g. tenrecs), Xenarthra, Euarchontoglires (e.g. primates and rodents) and Laurasiatheria (e.g. carnivores) (Madsen et al. 2001; Murphy et al. 2001 a, b).

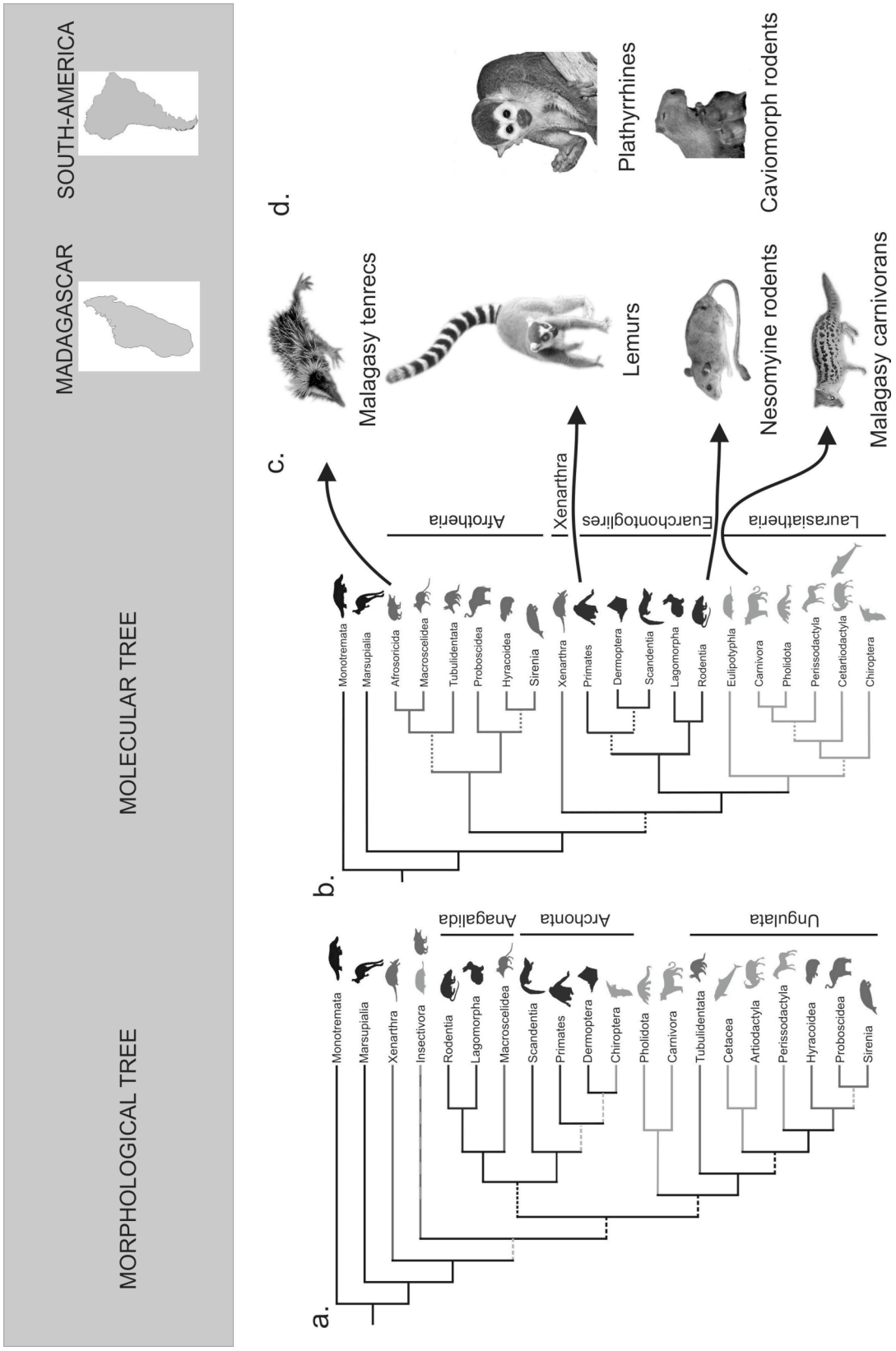

Figure 8: Mammalian phylogeny. (a) Placental mammal tree based on morphological and fossil evidence (ref), (b) placental tree based on molecular data (ref). The now generally accepted placental super-ordinal clades are: Xenarthra (sloths and armadillos, for example), Afrotheria (elephants, sea cows), Euarchontoglires (primates, bats, rodents) and Laurasiatheria (whales, carnivores, shrews). (c) The four terrestrial endemic mammalian clades of Madagascar: tenrecs, lemurs, rodents and carnivores, and (d) the two Paleogene endemic mammalian clades of South America: primates and rodents.

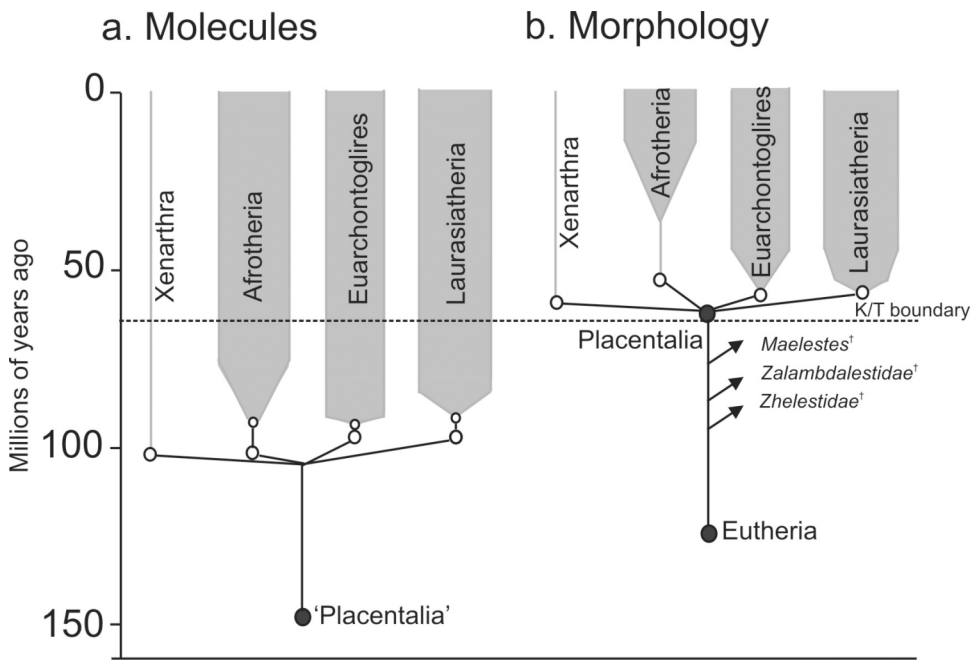

Figure 9: Schematic representations of the diversification of the major clades of placental mammals. (a) In the picture provided by the molecular analyses of Bininda-Emonds et al. (2007) the divergence of the four placental superorders occurred in the mid-Cretaceous, with ordinal diversification happening soon thereafter (although this is not the case for all lineages). (b) In the picture arising from the morphological (fossil) studies of Wible et al. (2007) the modern orders of placentals did not appear and diversify until after the K/T boundary. The fossils near the base of the tree are included in the broader group Eutheria, whose living representatives are the placentals. †, extinct group. Modified from Cifelli & Gordon (2007)

The eutherian divergence times have been subject to quite some debate between paleontologists and molecular systematists, and the most recent studies do not seem to narrow the gap between the conclusions drawn from morphological and molecular data (Fig.9). Following the extinction of non-avian dinosaurs at the K/T boundary, the fossil evidence shows that eutherians underwent significant "explosive" radiation in the Paleocene (65-55 MYA) (see Fig.10a for a geological timescale), and that most of the modern groups appeared and flourished later. This is the point of view of Wible et al. (2007) who consider the cretaceous eutherian fossils near the base of the mammalian tree as part of the stem group Eutheria and therefore not as direct ancestors of extant placentals (Fig.9). However, molecular studies have suggested much earlier divergence times of the eutherians: the four main placental superorders originated in the mid-cretaceous, (105 MYA (Springer et al. 2003; Bininda-Emonds et al. 2007), with the majority of intra-ordinal diversification happening soon thereafter. In this case the early diversification of mammals would rather be due to the break up of Gondwanaland (the Southern Hemisphere

supercontinent comprising Africa, South America, Antarctica, Australia, India and Madagascar; see Fig.10b for the evolution of paleo-costlines) rather than the extinction of dinosaurs (Eizirik et al. 2001). Both molecular and paleontological data are not exempt of pitfalls (see sections above) and congruence is likely to increase when paleontologists will fill the gaps in the fossil record and molecular systematists develop more sophisticated methods to account for rate variations.

Madagascar is one of the world's hottest biodiversity hot spots due to its diverse, endemic, and highly threatened biota. Only four lineages of extant, strictly terrestrial mammals are endemic to Madagascar: the tenrecs (Order Afrosoricida), the lemurs (Order Primates), the nesomyine rodents, and the carnivorans. Madagascar also supports a notable bat fauna, of which only the Myzopodidae family (the sucker-footed bats) is endemic. Relevant fossils are absent from Madagascar for the whole of the Tertiary period, and the rich findings from the Late Cretaceous include gondwanatheres, multituberculates and marsupials, but no fossils related to extant taxa (Krause et al. 1997a, b; Krause 2001). The extant mammal groups probably arrived during the Cenozoic (65-0 MYA) after the complete isolation of Madagascar (Krause et al. 1997a). The lack of relevant fossils from Madagascar leaves molecular studies as most promising to solve the question of the colonization of Madagascar after it became an island. The continental mass of Madagascar and India, broke away from Africa ((145 MYA) and began moving southeast, attaining its present position in front of Mozambique in the early Cretaceous, 130-118 MYA (Rabinowitz et al. 1983; Harland et al. 1990; Seward et al. 2004). Soon after their final separation from Africa, Antarctica and Australia began their southward movement away from IndoMadagascar (Briggs 2003), thus implying that terrestrial biotic exchange between IndoMadagascar and the remainder of Gondwana would have been impossible after 130 to 125 MYA. However, new fossil discoveries indicate a significant degree of cosmopolitanism among southern Gondwanan biota suggesting a contact between Antarctica and South America in the west, and between Antarctica and Indo-Madagascar in the east, apparently existing until circa 80 MYA (Krause et al. 1997a, b; Sampson et al. 1998, 2001; Buckley & Brochu 1999; Buckley et al. 2000). Finally, India separated from Madagascar in the late Cretaceous (100–80 MYA). A land bridge has been proposed (from ~45 to ~26 MYA) to have connected Africa and Madagascar (McCall 1997) and might have been a potential colonization route.

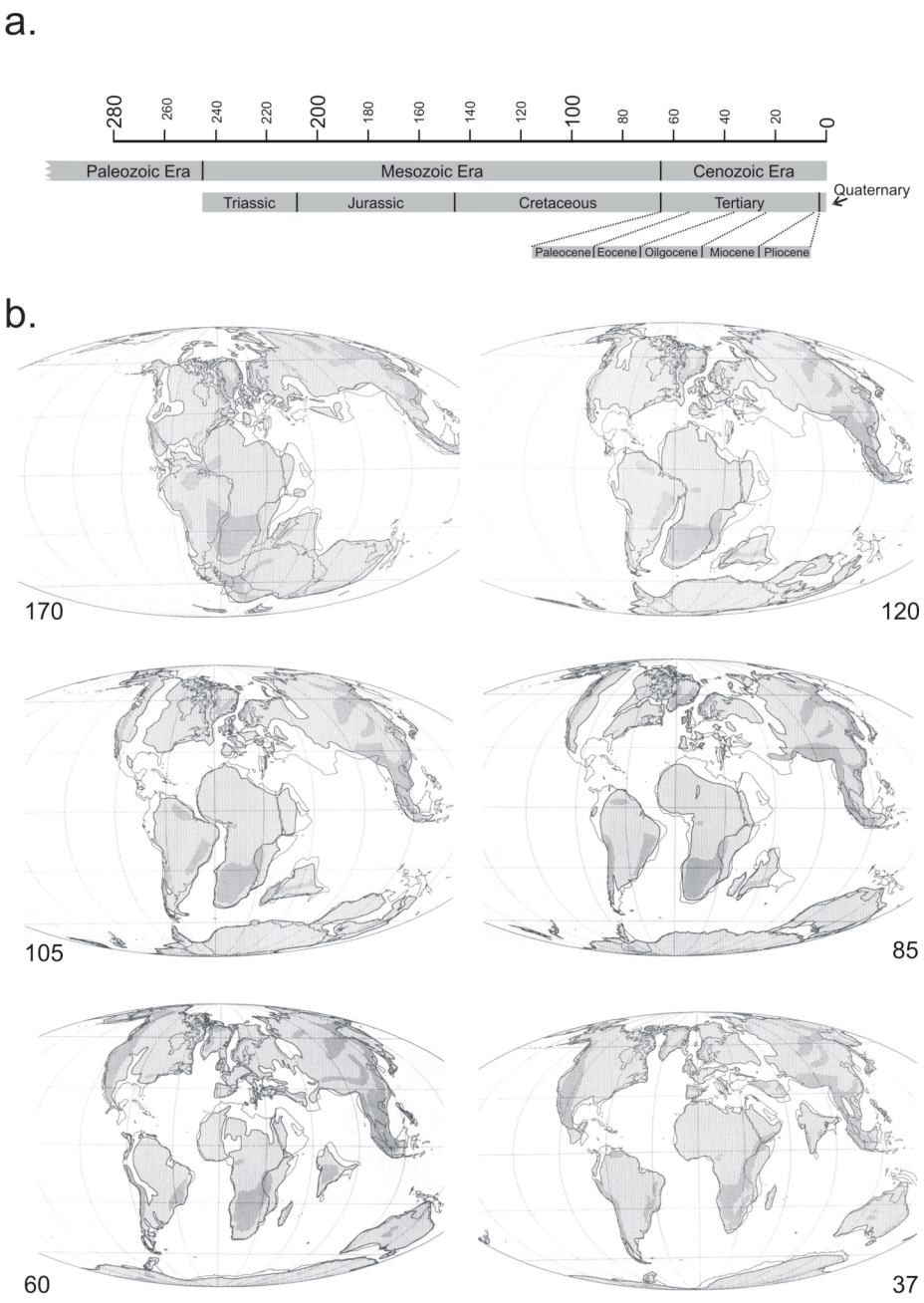

Figure 10: Geological timescale and Paleo-costline maps. (a) Geological timescale (in million years). (b) Paleo-costline maps at the indicated periods (in million years ago). Gray areas depict the paleo-continents, whereas present-day coastlines are indicated by lines. Maps taken from Smith et al. (1984).

South American mammalian evolution can be summarized by three successive phases. The oldest endemic fauna corresponds to the continent's early phase of isolation, and the establishment of basal clades within many of its indigenous, ancient lineages (e.g. xenarthrans, notoungulate and litoptern ungulates, and various marsupials). Only xenarthran and marsupial species are still extant in South America nowadays. This early period (~ 100 – 40 MYA) was dominated by

warm and humid tropical–temperate forest environments throughout the continent. The arrival of immigrant taxa, the caviomorph rodents and platyrrhine primates, and changes in the faunal aspects during the mid-Cenozoic reflecting adaptations such as hypsodont teeth to major environmental changes, including increased aridity and cooling, marks the base of the new phase. The last phase corresponds to the development of the recent Holocene fauna, culminating in the Great American Biotic Interchange (~ 4-3.5 MYA to the present) when the Isthmus of Panama rose up and bridged South and North America (Flynn & Wyss 1998). The fossil record of South America indicates that the new world mammalian biodiversity was higher in the past than it is today; thus many species have gone extinct owing to climate change, habitat fragmentation, disease and, more recently, human impact (MacFadden 2006). Even though the Paleogene South American fossil record is sparse, it can tell us that primates and rodents were already present in South America in the Early Eocene (28.4-33.9 MYA). Concerning the tectonic plate movements the following account is a synthesis of ideas presented in Sanmartín & Ronquist (2004) and Upchurch (2008). South America began to separate from Africa in the Early Cretaceous (135-105 MYA) with the opening of the South Atlantic Ocean at the latitude of Argentina and Chile. Northern South America and Africa remained connected until the mid-late Cretaceous (110–95 MYA), when a transform fault opened between Brazil and Guinea. As a result, South America drifted southwest into contact with Antarctica. Therefore Australia and South America remained in contact through Antarctica until the Eocene, when Antarctica completely separated from Australia (35 MYA) with the opening of the South Tasman Sea. On the West side, South America and Antarctica remained in contact through the Antarctic Peninsula until the Oligocene (30–28 MYA), when the Drake Passage opened between these continents, allowing the establishment of the Antarctic Circumpolar Current and the onset of the first Antarctic glaciations. Although landmass connections persisted until the Oligocene, the sparse Antarctic fauna seems endemic compared with contemporary high-latitude South American faunas (low levels of taxonomic similarity) (Woodburne & Case 1996), suggesting that biotic connections were disrupted before the Late Eocene (perhaps by 40–55 MYA).

Refined knowledge of tectonic plate movements is crucial when one aims at the reconstruction of biogeographic scenarios. Indeed, as mentioned by Upchurch (2008), the possible disconnection / reconnection periods between landmasses can

affect the organismal distribution and blur biogeographic signals. As a result, distinguishing between vicariance and dispersal to explain distribution patterns can be a difficult task. That is perhaps why making such distinctions remains one of the most debated biogeographic questions concerning the southern hemisphere.

Biogeographic scenarios: vicariance vs. dispersal

A classic problem in biogeography is to explain why particular terrestrial and freshwater taxa have geographical distributions that are broken up by oceans (e.g. the southern beeches of the genus *Nothofagus* present in Australia, New Zealand, New Guinea and southern South America). From Darwin's time until the 1960s, the predominant answer to such questions was 'oceanic dispersal' (Nelson 1978). Although successful long-distance colonization was rarely witnessed, plausible dispersal mechanisms were easy to imagine (e.g. plant seeds). More significantly, proponents of oceanic dispersal argued that some islands, like Mauritius and Hawaii, had never been connected to other landmasses; thus, the ancestors of all native organisms on such islands must have arrived by over-water dispersal. During the 1960s and 1970s, two developments started a revolution in historical biogeography that drastically reduced the perceived importance of oceanic dispersal (Wiley 1988). The first was the validation of the plate-tectonics theory that provided vicariance as explanation for biogeographical patterns on a global scale: widely distributed taxa on the ancient continent of Gondwana could have simply 'drifted' with the separating fragments of the continent. The second important development was the spreading of cladistic thinking. Cladistics provided an objective method for reconstructing phylogenetic relationships, and thus a mean to evaluate whether different taxa show histories of connections between areas that are concordant with each other and with the hypothesized vicariance history. The vicariance biogeography that emerged from the melding of plate tectonics and cladistics provided unifying explanations for the disjunct distributions of many taxa.

At this point dispersal was of course conceded for oceanic islands, but, for cases that could be explained by either vicariance or dispersal, most biogeographers assumed that vicariance was the more probable explanation as dispersal often was considered as almost impossible (Wiley 1988). However, the past few years have seen a strong increase in the number of studies supporting oceanic dispersal, and often in

cases that had been explained previously by vicariance (e.g. Baum et al. 1998; Givnish et al. 2000; Trewick 2000; Renner et al. 2000; Vences et al. 2001, 2003; Davis et al. 2002; McDowall 2002; Raxworthy at al. 2002; Vicario et al. 2003). Collectively, these studies represent a major shift in historical biogeography that has profound implications both for how we view the geographical history of biotas and for the methods that we use to decipher that history. In a broad analysis of Southern Hemisphere taxa, Sanmartín & Ronquist (2004) found, in agreement with other studies, that the animal data are congruent with the geological sequence of Gondwana breakup. Trans-Antarctic dispersal is also significantly more frequent than any other dispersal event in animals, which may be explained by the long period of geological contact between Australia and South America via Antarctica. In contrast, the dominant pattern in plants is better explained by dispersal. This is consistent with the notion that animals have more difficulty in crossing ocean barriers than do plants. However, the debate is not over; Upchurch (2008) recently emphasized that in addition to vicariance and trans-oceanic dispersal, organismal distributions are affected by extinctions, sampling errors and geodispersal (i.e. the expansion of species ranges to new areas after a geographical barrier has been removed), and only few studies of Gondwanan biogeography have examined the effect of the latter two on the accuracy of their conclusions.

Evolution on islands

Biota on islands show a distinct signature of evolution in isolation, reflected by both the imbalance presence of the lineages and the high levels of diversity within lineages.

Oceanic islands are difficult to reach, and since organisms of different taxa have different dispersal abilities it is inevitable that these islands will possess a non-representative sample of species from the nearest mainland. Moreover, chance plays a strong role in determining which species arrive on a given island, when, and in what numbers. Difficulties of establishment in the new environment will further influence the composition of the island community by favoring some types of colonists over others. All these parameters taken together may explain why mammals are usually underrepresented on islands. The evolution of the colonizer will then take place in a physically and biotically novel environment.

It is generally agreed that colonization of large islands is followed by adaptive radiation (i.e. the rapid proliferation of an ecologically and morphologically differentiated species assemblage from one ancestral species as a consequence of the adaptation to various ecological niches), which is thought to have played a prominent role in organismal diversification (Schluter 2000). The spectacular diversity of forms that have emerged in adaptive radiations, and the explosive mode of species formation, have fascinated empiricists and theoreticians alike: adaptive radiation can result from many processes, e.g. from ecological speciation as it was shown for the "Darwin's finches" from the Galapagos islands (Ryan et al. 2007) to speciation triggered by sexual selection (Stelkens et al. 2008) or hybridization (reviewed in Bell & Travis 2005) as demonstrated for the cichlid fishes from the African great lakes.

Morphologically speaking, the tenrec radiation on Madagascar seems to result from an adaptive radiation. Moreover, species-level phylogenies derived from molecular data provide an indirect record of the speciation events that have led to extant species. This offers enormous potential for investigating the general causes and rates of speciation within clades. To make the most of this potential, one should ideally sample all the species in a higher group, such as a genus, ensure that those species reflect evolutionary entities within the group, and rule out the effects of other processes, such as extinction, as explanation for the observed patterns (Barraglough & Nee 2001). However, the ideal sampling is often difficult to reach, especially for large clades, due to both sequencing costs (nowadays becoming cheaper) and difficulties in the field to catch the required specimens.

AIMS AND OUTLINES OF THE THESIS

The research described in this thesis focuses on the assessment of the biogeographical history of various endemic mammalian groups from South America and Madagascar. Trying to reconstruct the evolutionary history of the fauna and the flora is an important question, as ultimately it can contribute to understanding how global changes (e.g. continental breakup, climatic changes) affect organismal distribution and diversity, and speciation processes.

To be able to reconstruct the biogeographical history of a group it is a first necessity to get a solid phylogeny. In chapter 2, using the phylogenetic information

provided by indels as Rare Genomic Changes (RGCs), we brought strong arguments in favor of the monophyly of the newly recognized superordinal mammalian clade Euarchontoglires. As already mentioned in the previous section, this grouping has been heavily debated amongst molecular phylogeneticists depending on whether they were using nuclear or mitochondrial data.

The third chapter focuses on primate phylogeny. This order contains two subfamilies that are endemic to Madagascar (the lemurs) and to South America (the platyrrhine primates). In this chapter datings have been done using local molecular clocks. The incompatibility as seen between some paleontological and molecular estimates may reflect the incompleteness of the fossil record, but may more surely indicate that the variability of evolutionary rates cannot be fully accommodated by local clock methods.

Consequently, in the following chapters we switched to the Bayesian relaxed clock methods. Chapter 4 and 5 assessed the timing of colonization and diversification of the endemic mammals of South America and Madagascar, respectively. The results showed that, in contrast to what was thought, mammals can repeatedly disperse through large oceanic barriers. The results of the analyses for each endemic mammalian group are strikingly uniform. For all of them, phylogenetic analyses demonstrated their monophyly, with their respective sister groups found in Africa. The colonizations of Madagascar were asynchronous; the picture was less clear for South America.

The last scientific chapter is focused on the evolution of Malagasy tenrecs. Their diversification pattern showed that morphological specializations of the tenrecs may have been affected by environmental changes caused by climatic and/or subsequent colonization events, and that major morphological specializations have appeared well after the adaptive radiation period.

Finally in chapter 7, the most important results and conclusions of this study are summarized and discussed in relation to recent developments in the field and with regard to their implications for further research in phylogeography, molecular dating and protein evolution.

REFERENCES

Aris-Brosou, S. (2007). Dating phylogenies with hybrid local molecular clocks. *PLoS ONE* 2: e879.

Aris-Brosou, S., and Z. Yang. (2002). Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Syst. Biol.* 51: 703-714.

Aris-Brosou, S., and Z. Yang. (2003). Bayesian models of episodic evolution support a late precambrian explosive diversification of the Metazoa. *Mol. Biol. Evol.* 20: 1947-1954.

Ayala, F. J. (1999). Molecular clock mirages. *Bioessays* 21: 71-75.

Bandelt, H.-J., Q.-P. Kong, M. Richards, and V. Macaulay. (2006). Estimation of mutation rates and coalescence times: some caveats *in* Human Mitochondrial DNA and the Evolution of Homo Sapiens, H.-J. Bandelt, V. Macaulay, and M. Richards, eds. Pp 47-90. Springer-Verlag, Berlin, Heidelberg.

Barraclough, T. G., and S. Nee. (2001). Phylogenetics and speciation. *Trends Ecol. Evol.* 16: 391-399.

Baum, D. A., R. L. Small, and J. F. Wendel. (1998). Biogeography and floral evolution of baobabs (Adansonia, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* 47: 181-207.

Bedford, T., and D. L. Hartl. (2008). Overdispersion of the molecular clock: temporal variation of gene-specific substitution rates in Drosophila. *Mol. Biol. Evol.* 25: 1631-1638.

Bell, M. A., and M. P. Travis. (2005). Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. *Trends Ecol Evol* 20: 358-361.

Benton, M. J., and F. J. Ayala. (2003). Dating the tree of life. *Science* 300: 1698-1700.

Benton, M. J., and P. C. Donoghue. (2007). Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 24: 26-53.

Bickel, D. R. (2000). Implications of fluctuations in substitution rates: impact on the uncertainty of branch lengths and on relative-rate tests. *J. Mol. Evol.* 50: 381-390.

Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. MacPhee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. (2007). The delayed rise of present-day mammals. *Nature* 446: 507-512.

Blair, J. E., and S. B. Hedges. (2005). Molecular clocks do not support the Cambrian explosion. *Mol. Biol. Evol.* 22: 387-390.

Blair, J. E., K. Ikeo, T. Gojobori, and S. B. Hedges. (2002). The evolutionary position of nematodes. *BMC Evol. Biol.* 2: 7.

Briggs, J. C. (2003). The biogeographic and tectonic history of India. *J. Biogeogr.* 30: 381-388.

Brinkmann, H., M. van der Giezen, Y. Zhou, G. Poncelin de Raucourt, and H. Philippe. (2005). An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol.* 54: 743-757.

Britten, R. J. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231: 1393-1398.

Bromham, L., and M. Cardillo. (2003). Testing the link between the latitudinal gradient in

species richness and rates of molecular evolution. *J. Evol. Biol.* 16: 200-207.

Bromham, L., and D. Penny. (2003). The modern molecular clock. *Nat. Rev. Genet.* 4: 216-224.

Bromham, L., D. Penny, A. Rambaut, and M. D. Hendy. (2000b). The power of relative rates tests depends on the data. *J. Mol. Evol.* 50: 296-301.

Bromham, L., M. J. Phillips, and D. Penny. (1999). Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends Ecol. Evol.* 14: 113-118.

Bromham, L., A. Rambaut, R. Fortey, A. Cooper, and D. Penny. (1998). Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl. Acad. Sci. USA* 95: 12386-12389.

Bromham, L., M. Woolfit, M. S. Lee, and A. Rambaut. (2002). Testing the relationship between morphological and molecular rates of change along phylogenies. *Evolution* 56: 1921-1930.

Bromham, L. D., and M. D. Hendy. (2000a). Can fast early rates reconcile molecular dates with the Cambrian explosion? *Proc. Biol. Sci.* 267: 1041-1047.

Buckley, G. A., and C. A. Brochu. (1999). An enigmatic new crocodile from the Upper Cretaceous of Madagascar *in* Cretaceous Fossil Vertebrates, D. Unwin, ed. Pp 147-155. Paleontological Association, London.

Buckley, G. A., C. A. Brochu, D. W. Krause, and D. Pol. (2000). A pug-nosed crocodyliform from the Late Cretaceous of Madagascar. *Nature* 405: 941-944.

Cifelli R.L., and C.L. Gordon. (2007). Re-crowning mammals. *Nature* 447: 918-920.

Claverie, J. M. (2005). Fewer genes, more noncoding RNA. *Science* 309: 1529-1530.

Consortium I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

Consortium I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.

Cooper, A., and D. Penny. (1997). Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* 275: 1109-1113.

Cutler, D. J. (2000). Estimating divergence times in the presence of an overdispersed molecular clock. *Mol. Biol. Evol.* 17: 1647-1660.

D'Erchia, A. M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. (1996). The guinea-pig is not a rodent. *Nature* 381: 597-600.

Davis, C. C., C. D. Bell, P. W. Fritsch, and S. Mathews. (2002). Phylogeny of Acridocarpus-Brachylophon (Malpighiaceae): implications for tertiary tropical floras and Afroasian biogeography. *Evolution* 56: 2395-2405.

Dickerson, R. E. (1971). The structures of cytochrome c and the rates of molecular evolution. *J. Mol. Evol.* 1: 26-45.

Donoghue, P. C., and M. J. Benton. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22: 424-431.

Doolittle, R. F., and B. Blombaeck. (1964). Amino-Acid Sequence Investigations of Fibrinopeptides from Various Mammals: Evolutionary Implications. *Nature* 202: 147-152.

Douzery, E. J., F. Delsuc, M. J. Stanhope, and D. Huchon. (2003). Local molecular clocks in

three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J. Mol. Evol.* 57: S201-S213.

Douzery, E. J., E. A. Snell, E. Bapteste, F. Delsuc, and H. Philippe. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. USA* 101: 15386-15391.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. (1998). Rates of spontaneous mutation. *Genetics* 148: 1667-1686.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4: e88.

Drummond, A. J., and A. Rambaut. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.

Easteal, S. (1985). Generation time and the rate of molecular evolution. *Mol. Biol. Evol.* 2: 450-453.

Easteal, S., C. Collet, and D. Betty. (1995). The mammalian molecular clock. Springer-Verlag, and R.G. Landes, New York, London, Austin.

Eizirik, E., W. J. Murphy, and S. J. O'Brien. (2001). Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* 92: 212-219.

Emerson, B. C. (2007). Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst. Biol.* 56: 337-345.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401-410.

Felsenstein, J., and G. A. Churchill. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13: 93-104.

Fitch, W. M. (1976). Molecular evolutionary clocks *in* Molecular evolution, F. J. Ayala, ed. Pp 160-178. Sinauer Associates, Sunderland.

Fitch, W. M., and E. Markowitz. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.* 4: 579-593.

Flynn, J. J., and A. R. Wyss. (1998). Recent advances in South American mammalian paleontology. *Trends Ecol. Evol.* 13: 449-454.

Foote, M., J. P. Hunter, C. M. Janis, and J. J. J. Sepkoski. (1999). Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherians mammals. *Science* 283: 1310-1314.

Foster, P. G. (2004). Modeling compositional heterogeneity. *Syst. Biol.* 53: 485-495.

Galtier, N. (2001). Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* 18: 866-873.

Galtier, N., and M. Gouy. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15: 871-879.

Garcıa-Moreno, J. (2004). Is there a universal mtDNA clock for birds? *J. Avian Biol.* 35: 465–

468.

Gilbert, W. (1978). Why genes in pieces? *Nature* 271: 501.

Gillespie, J. H. (1991). The causes of molecular evolution. Oxford University Press, New York.

Givnish, T. J., T. M. Evans, M. L. Zjhra, T. B. Patterson, P. E. Berry, and K. J. Sytsma. (2000). Molecular evolution, adaptive radiation, and geographic diversification in the amphiatlantic family Rapateaceae: evidence from ndhF sequences and morphology. *Evolution* 54: 1915-1937.

Goodman, M. (1985). Rates of molecular evolution: the hominoid slowdown. *BioEssays* 3: 9-14.

Goremykin, V. V., K. I. Hirsch-Ernst, S. Wolfl, and F. H. Hellwig. (2003). Analysis of the Amborella trichopoda chloroplast genome sequence suggests that amborella is not a basal angiosperm. *Mol. Biol. Evol.* 20: 1499-505.

Graur, D., and W. Martin. (2004). Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics* 20: 80-86.

Gregory, T. R. (2005). Synergy between sequence and size in large-scale genomics. *Nat. Rev. Genet.* 6: 699-708.

Gregory, T. R., and R. DeSalle. (2005). Comparative genomics in prokaryotes *in* The Evolution of the Genome, T. R. Gregory, ed. Pp 585-675. Elsevier, San Diego.

Harland, W. B., R. L. Amstrong, A. V. Cox, L. E. Craiq, A. G. Smith, and D. G. Smith. (1990). A Geological Time Scale. Cambridge University Press, Cambridge.

Heads, M. (2005). Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* 21: 62-78.

Hedges, S. B., and S. Kumar. (2003). Genomic clocks and evolutionary timescales. *Trends Genet.* 19: 200-206.

Hedges, S. B., and S. Kumar. (2004). Precision of molecular time estimates. *Trends Genet.* 20: 242-247.

Hendy, M. D., and D. Penny. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38: 297-309.

Ho, S. Y. (2007). Calibrating molecular estimates of substitution rates and divergence times in birds. *J. Avian Biol.* 38: 409-414.

Ho, S. Y., and G. Larson. (2006). Molecular clocks: when times are a-changin'. *Trends Genet.* 22: 79-83.

Ho, S. Y., M. J. Phillips, A. Cooper, and A. J. Drummond. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22: 1561-1568.

Ho, S. Y., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. (2007). Evidence for time dependency of molecular rate estimates. *Syst. Biol.* 56: 515-522.

Holder, M., and P. O. Lewis. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4: 275-284.

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull, and C. Herrnstadt. (2003). The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.* 72: 659-670.

Huchon, D., O. Madsen, M. J. Sibbald, K. Ament, M. J. Stanhope, F. Catzeflis, W. W. de Jong, and E. J. Douzery. (2002). Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* 19: 1053-1065.

Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51: 673-688.

Huelsenbeck, J. P., B. Larget, and D. Swofford. (2000). A compound poisson process for relaxing the molecular clock. *Genetics* 154: 1879-1892.

Inoue, J., P. C. J. Donoghue, and Z. Yang. (Submitted). The impact of the representation of fossil calibrations on Baysian Estimation of species divergence times.

Janke, A., O. Magnell, G. Wieczorek, M. Westerman, and U. Arnason. (2002). Phylogenetic Analysis of 18S rRNA and the Mitochondrial Genomes of the Wombat, *Vombatus ursinus*, and the Spiny Anteater, *Tachyglossus aculeatus*: Increased Support for the Marsupionta Hypothesis. *J. Mol. Evol.* 54: 71-80.

Killian, J. K., T. R. Buckley, N. Stewart, B. L. Munday, and R. L. Jirtle. (2001). Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. *Mamm. Genome* 12: 513-517.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* 217: 624-626.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

King, J. L., and T. H. Jukes. (1969). Non-Darwinian evolution. *Science* 164: 788-798.

Kishino, H., and M. Hasegawa. (1990). Converting distance to time: application to human evolution. *Methods Enzymol.* 183: 550-570.

Kishino, H., J. L. Thorne, and W. J. Bruno. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* 18: 352-361.

Kohne, D. E. (1970). Evolution of higher-organism DNA. *Q. Rev. Biophys.* 3: 327-375.

Koshi, J. M., and R. A. Goldstein. (2001). Analyzing site heterogeneity during protein evolution. *Pac. Symp. Biocomput.*: 191-202.

Krause, D. W. (2001). Fossil molar from a Madagascan marsupial. *Nature* 412: 497-8.

Krause, D. W., J. H. Hartman, and N. A. Wells. (1997a). Late Cretaceous vertebrates from Madagascar. Implications for biotic changes in deep time *in* Natural change and human impact in Madagascar, S. M. G. B. D. Patterson, ed. Pp. 3-43. Smithsonian Institution Press, Washington.

Krause, D. W., G. V. R. Prasad, W. von Koenigswald, A. Sahni, and F. E. Grine. (1997b). Cosmopolitanism among Gondwanan Late Cretaceous mammals. *Nature* 390: 504-507.

Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* 6: 654-662.

Kumar, S., and S. B. Hedges. (1998). A molecular timescale for vertebrate evolution. *Nature* 392: 917-920.

Laird, C. D., B. L. Mac Conaughy, and B. J. Mac Carthy. (1969). Rate of fixation of nucleotide substitutions in evolution. *Nature* 224: 149-154.

Lambert, D. M., P. A. Ritchie, C. D. Millar, B. Holland, A. J. Drummond, and C. Baroni. (2002). Rates of evolution in ancient DNA from Adelie penguins. *Science* 295: 2270-2273.

Lander, E. S., L. M. Linton, B. Birren, et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Langley, C. H., and W. M. Fitch. (1974). An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* 3: 161-177.

Lartillot, N., H. Brinkmann, and H. Philippe. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* 7 Suppl 1: S4.

Lartillot, N., and H. Philippe. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21: 1095-1109.

Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. (2007). A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* 24: 2669-2680.

Li, W.-H. (1993). So, what about the molecular clock hypothesis? *Curr. Opin. Genet. Develop.* 3: 896-901.

Li, W.-H. (1997). Molecular Evolution. Sinauer Associates, Sunderland.

Lopez, P., D. Casane, and H. Philippe. (2002). Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19: 1-7.

Macfadden, B. J. (2006). Extinct mammalian biodiversity of the ancient New World tropics. *Trends Ecol. Evol.* 21: 157-165.

Madsen, O., M. Scally, C. J. Douady, D. J. Kao, R. W. DeBry, R. Adkins, H. M. Amrine, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409: 610-4.

Margoliash, E. (1963). Primary Structure and Evolution of Cytochrome C. *Proc. Natl. Acad. Sci. USA* 50: 672-679.

Margoliash, E., and S. E. L. (1965). Structural and functional aspects of cytochrome c in relation to evolution *in* Evolving genes and proteins, V. Bryson, and H. J. Vogel, eds. Pp 221-242. Academic Press, New York.

Martin, A. P., and S. R. Palumbi. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. USA* 90: 4087-4091.

McCall, R. A. (1997). Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc. R. Soc. Lond. B Biol. Sci.* 264: 663-665.

McDowall, R. M. (2002). Accumulating evidence for a dispersal biogeography of southern cool temperate freshwater fishes. *J. Biogeogr.*: 207-219.

Muller, J., and R. R. Reisz. (2005). Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* 27: 1069-1075.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001a). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294: 2348-2351.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. (2001b). Molecular phylogenetics and the origins of placental mammals. *Nature* 409: 614-8.

Murphy , W. J., T.H. Pringle, T.A. Crider, M. S. Springer, and W. Miller. (2007). Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* 17: 413-421.

Near, T. J., and M. J. Sanderson. (2004). Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359: 1477-1483.

Nei, M., P. Xu, and G. Glazko. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. USA* 98: 2497-2502.

Nelson, G. (1978). From Candolle to croizat: Comments on the history of biogeography. *J. Hist. Biol.* 11: 269-305.

Novacek, M. J. (1992). Mammalian phylogeny: shaking the tree. *Nature* 356: 121-125.

Ohno, S. (1972). So much 'junk' data in our genome *in* Evolution of Genetic Systems, H. H. Smith, ed. Pp 366–370. Gordon-Breach, New York.

Ohta, T. (1972). Population size and rate of evolution. *J. Mol. Evol.* 1: 305-314.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96-98.

Ohta, T. (1987). Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* 26: 1-6.

Ohta, T. (2002). Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. USA* 99: 16134-16137.

Ohta, T. (2003). Origin of the neutral and nearly neutral theories of evolution. *J. Biosci.* 28: 371-377.

Ohta, T., and M. Kimura. (1971). On the constancy of the evolutionary
rate of cistrons. *J. Mol. Evol.* 1: 18–25.

Pagel, M., and A. Meade. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53: 571-581.

Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. USA* 96: 3807-3812.

Penny, D. (2005). Evolutionary biology: relativity for molecular clocks. *Nature* 436: 183-184.

Penny, D., B. J. McComish, M. A. Charleston, and M. D. Hendy. (2001). Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* 53: 711-723.

Perez-Losada, M., J. T. Hoeg, and K. A. Crandall. (2004). Unraveling the evolutionary radiation of the thoracican barnacles using molecular and morphological evidence: a comparison of several divergence time estimation approaches. *Syst. Biol.* 53: 244-264.

Peterson, K. J., J. B. Lyons, K. S. Nowak, C. M. Takacs, M. J. Wargo, and M. A. McPeek. (2004). Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. USA* 101: 6536-6541.

Philippe, H., and A. Germot. (2000). Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.* 17: 830-834.

Philippe, H., N. Lartillot, and H. Brinkmann. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* 22:

1246-1253.

Philippe, H., and J. Laurent. (1998). How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* 8: 616-623.

Philippe, H., P. Lopez, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Muller, and H. Le Guyader. (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. Biol. Sci.* 267: 1213-1221.

Poux, C., P. Chevret, D. Huchon, W. W. de Jong, and E. J. Douzery. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.* 55: 228-244.

Poux, C., O. Madsen, J. Glos, W. W. de Jong, and M. Vences. (2008). Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses. *BMC Evol. Biol.* 8: 102.

Poux, C., O. Madsen, E. Marquard, D. R. Vieites, W. W. de Jong, and M. Vences. (2005). Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst. Biol.* 54: 719-730.

Pulquerio, M. J., and R. A. Nichols. (2007). Dates from the molecular clock: how wrong can we be? *Trends Ecol. Evol.* 22: 180-184.

Rabinowitz, P. D., M. F. Coffin, and D. Falvey. (1983). The separation of Madagascar and Africa. *Science* 220: 67-69.

Rambaut, A., and L. Bromham. (1998). Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15: 442-448.

Rand, D. M. (1994). Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends Ecol. Evol.* 9: 125–131.

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51: 754-760.

Rannala, B., and Z. Yang. (2007). Inferring speciation times under an episodic molecular clock. *Syst. Biol.* 56: 453-466.

Raxworthy, C. J., M. R. Forstner, and R. A. Nussbaum. (2002). Chameleon radiation by oceanic dispersal. *Nature* 415: 784-787.

Reisz, R. R., and J. Muller. (2004). Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* 20: 237-241.

Renner, S. S., D. B. Foreman, and D. Murray. (2000). Timing transantarctic disjunctions in the Atherospermataceae (Laurales): evidence from coding and noncoding chloroplast sequences. *Syst. Biol.* 49: 579-591.

Reyes, A., C. Gissi, G. Pesole, F. M. Catzeflis, and C. Saccone. (2000). Where Do Rodents Fit? Evidence from the Complete Mitochondrial Genome of Sciurus vulgaris. *Mol. Biol. Evol.* 17: 979-983.

Robinson, D. M., D. T. Jones, H. Kishino, N. Goldman, and J. L. Thorne. (2003). Protein evolution with dependence among codons due to tertiary structure. *Mol. Biol. Evol.* 20: 1692-1704.

Robinson, M., M. Gouy, C. Gautier, and D. Mouchiroud. (1998). Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* 15: 1091-1098.

Rodrigue, N., N. Lartillot, D. Bryant, and H. Philippe. (2005). Site interdependence attributed to tertiary structure in amino acid sequence evolution. *Gene* 347: 207-217.

Rokas, A., and P. W. Holland. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15: 454-459.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798-804.

Rutschmann, F., T. Eriksson, K. A. Salim, and E. Conti. (2007). Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. *Syst. Biol.* 56: 591-608.

Ryan, P. G., P. Bloomer, C. L. Moloney, T. J. Grant, and W. Delport. (2007). Ecological speciation in South Atlantic island finches. *Science* 315: 1420-1423.

Sampson, S. D., M. T. Carrano, and C. A. Forster. (2001). A bizarre predatory dinosaur from the Late Cretaceous of Madagascar. *Nature* 409: 504-506.

Sampson, S. D., L. M. Witmer, C. A. Forster, D. W. Krause, P. M. O'Connor, P. Dodson, and F. Ravoavy. (1998). Predatory dinosaur remains from madagascar: implications for the cretaceous biogeography of gondwana. *Science* 280: 1048-1051.

Sanders, K. L., and M. S. Y. Lee. (2007). Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biol. Lett.* 3: 275-279.

Sanderson, M. J. (1997). A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy. *Mol. Biol. Evol.* 14: 1218-1231.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19: 101-109.

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301-302.

Sanger, F. (1959). Chemistry of insulin; determination of the structure of insulin opens the way to greater understanding of life processes. *Science* 129: 1340-1344.

Sanmartin, I., and F. Ronquist. (2004). Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst. Biol.* 53: 216-243.

Sarich, V. M., and A. C. Wilson. (1973). Generation time and genomic evolution in primates. *Science* 179: 1144-1147.

Scherer, S. (1989). The relative-rate test of the molecular clock hypothesis: a note of caution. *Mol. Biol. Evol.* 6: 436-441.

Schluter, D. (2000). The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.

Seward, D., D. Grujic, and G. Schreurs. (2004). An insight into the breakup of Gondwana: Identifying events through low-temperature thermochronology from the basement rocks of Madagascar. *Tectonics* 23: C3007.

Shabalina S.A., and N.A. Spiridonov. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. *Genome Biol.* 5: 105.

Shaul, S., and D. Graur. (2002). Playing chicken (Gallus gallus): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene* 300: 59-61.

Siepel, A., and D. Haussler. (2004). Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* 21: 468-488.

Simpson, A. G., and A. J. Roger. (2004). The real 'kingdoms' of eukaryotes. *Curr. Biol.* 14: R693-R696.

Smith, A. B., and K. J. Peterson. (2002). Dating the time of origin of major clades: molecular clocks and the fossil record. *Ann. Rev. Earth Planet. Sci.* 30: 65-88.

Smith, A. B., D. Pisani, J. A. Mackenzie-Dodds, B. Stockley, B. L. Webster, and D. T. Littlewood. (2006). Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol. Biol. Evol.* 23: 1832-1851.

Smith, A.G., D.G. Smith, and B.M. Funell. (1994). Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.

Soltis, D. E., V. A. Albert, V. Savolainen, K. Hilu, Y. L. Qiu, M. W. Chase, J. S. Farris, S. Stefanovic, D. W. Rice, J. D. Palmer, and P. S. Soltis. (2004). Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. *Trends Plant. Sci.* 9: 477-483.

Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. (2002). Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. USA* 99: 4430-4435.

Springer, M. S. (1997). Molecular clocks and the timing of the placental and marsupial radiations in relation to the Cretaceous-Tertiary boundary. *J. Mammal. Evol.* 4: 285-302.

Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope. (2001). Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18: 132-143.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100: 1056-1061.

Stefanovic, S., D. W. Rice, and J. D. Palmer. (2004). Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? *BMC Evol. Biol.* 4: 35.

Stelkens, R. B., M. E. Pierotti, D. A. Joyce, A. M. Smith, I. van der Sluijs, and O. Seehausen. (2008). Disruptive sexual selection on male nuptial coloration in an experimental hybrid population of cichlid fish. *Philos Trans R Soc Lond B Biol Sci* 363: 2861-2870.

Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135: 599-607.

Takezaki, N., A. Rzhetsky, and M. Nei. (1995). Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12: 823-833.

Tavare, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. (2002). Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* 416: 726-729.

Thorne, J. L., and H. Kishino. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51: 689-702.

Thorne, J. L., H. Kishino, and I. S. Painter. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15: 1647-1657.

Torrents, D., M. Suyama, E. Zdobnov, and P. Bork. (2003). A genome-wide survey of human pseudogenes. *Genome Res*. 13: 2559-2567.

Trewick, S. A. (2000). Molecular evidence for dispersal rather than vicariance as the origin of flightless insect species on the Chatham Islands, New Zealand. *J. Biogeogr.* 27: 1189-1200.

Tuffley, C., and M. Steel. (1998). Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci.* 147: 63-91.

Upchurch, P. (2008). Gondwanan break-up: legacies of a lost world? *Trends Ecol Evol* 23: 229-236.

van Tuinen, M., and G. J. Dyke. (2004). Calibration of galliform molecular clocks using multiple fossils and genetic partitions. *Mol. Phylogenet. Evol.* 30: 74-86.

van Tuinen, M., and E. A. Hadly. (2004a). Error in estimation of rate and time inferred from the early amniote fossil record and avian molecular clocks. *J. Mol. Evol.* 59: 267-276.

van Tuinen, M., and E. A. Hadly. (2004b). Calibration and error in placental molecular clocks: a conservative approach using the cetartiodactyl fossil record. *J. Hered.* 95: 200-208.

van Tuinen, M., and S. B. Hedges. (2001). Calibration of avian molecular clocks. *Mol. Biol. Evol.* 18: 206-213.

Vences, M., J. Freyhof, R. Sonnenberg, J. Kosuch, and M. Veith. (2001). Reconciling fossils and molecules: Cenozoic divergence of cichlid fishes and the biogeography of Madagascar. *J. Biogeogr.* 28: 1091-1099.

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer. (2003). Multiple overseas dispersal in amphibians. *Proc R Soc Lond B Biol Sci* 270: 2435-42.

Vicario, S., A. Caccone, and J. Gauthier. (2003). Xantusiid "night" lizards: a puzzling phylogenetic problem revisited using likelihood-based Bayesian methods on mtDNA sequences. *Mol. Phylogenet. Evol.* 26: 243-261.

Welch, J. J., E. Fontanillas, and L. Bromham. (2005). Molecular dates for the "cambrian explosion": the influence of prior assumptions. *Syst. Biol.* 54: 672-678.

Wible, J. R., G. W. Rougier, M. J. Novacek, and R. J. Asher. (2007). Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. *Nature* 447: 1003-1006.

Wiley, E. O. (1988). Vicariance Biogeography. *Annu. Rev. Ecol. Syst.* 19: 513-542.

Wolf, Y. I., I. B. Rogozin, and E. V. Koonin. (2004). Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* 14: 29-36.

Woodburne, M. O., and J. A. Case. (1996). Dispersal, vicariance, and the Late Cretaceous to early tertiary land mammal biogeography from South America to Australia. *J. Mammal. Evol.* 3: 121-161.

Wray, G. A., J. S. Levinton, and L. H. Shapiro. (1996). Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla. *Science* 274: 568-573.

Wu, C.-I., and W.-H. Li. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* 82: 1741-1745.

Wyles, J. S., J. G. Kunkel, and A. C. Wilson. (1983). Birds, behavior, and anatomical evolution. *Proc. Natl. Acad. Sci. USA* 80: 4394-4397.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39: 306-314.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11: 367-372.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13: 555-556.

Yang, Z. (2004). A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica* 50: 645-656.

Yang, Z., and B. Rannala. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23: 212-226.

Yoder, A. D., and Z. Yang. (2000). Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17: 1081-1090.

Yoder, A. D., and Z. Yang. (2004). Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* 13: 757-773.

Zhang, J., Y. P. Zhang, and H. F. Rosenberg. (2002). Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* 30: 411-415.

Zuckerkandl, E., and L. Pauling. (1965). Evolutionary divergence and convergence in proteins *in* Evolving genes and proteins, V. Bryson, and H. J. Vogel, eds. Pp 97-166. Academic Press, New York.

Zuckerkandl, E., and L. B. Pauling. (1962). Molecular disease, evolution, and genetic heterogeneity *in* Horizons in Biochemistry, M. Kasha, and B. Pullman, eds. Pp 189–225. Academic Press, New York.

Chapter 2

Sequence Gaps Join Mice and Men: Phylogenetic Evidence from Deletions in Two Proteins

Céline Poux[1], Teun van Rheede[1], Ole Madsen[1], and Wilfried W. de Jong[1,2]

[1] Department of Biochemistry, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

Mol. Biol. Evol. 19:2035–2037. 2002

Recent nuclear sequence analyses have provided evidence that primates and rodents are more closely related than previously believed (Madsen et al. 2001; Murphy et al. 2001a, 2001b). This proposal is difficult to reconcile with morphological insights (Liu et al. 2001; Novacek 2001) and is not generally supported by current mitochondrial sequence data (Reyes et al. 2000; Nikaido et al. 2001; Arnason et al. 2002; Janke et al. 2002). Moreover, the supporting data and analyses have been criticized on methodological grounds (Rosenberg & Kumar 2001). Here we report deletions in two nuclear protein-coding genes that lend independent support to this contested grouping.

Some 18 orders of placental mammals are currently recognized, but their phylogenetic relationships remain highly controversial. Extensive sequence comparisons of mainly nuclear genes support a basal division into four major clades (Xenarthra, Afrotheria, Laurasiatheria, and Euarchontoglires), which has far-reaching implications for early mammalian biogeography and morphological diversification (Murphy et al. 2001b). Euarchontoglires is composed of the orders Primates, Rodentia, Lagomorpha (rabbits, hares, and pikas), Scandentia (tree shrews), and Dermoptera (flying lemurs). In contrast, morphology groups Primates, Scandentia, and Dermoptera with Chiroptera (bats) in the clade Archonta, whereas Rodentia and Lagomorpha (jointly called Glires) are in a distant clade with Macroscelidea (elephant shrews) (Liu et al. 2001; Novacek 2001). Also, sequence data from 12 proteins encoded by the mitochondrial genome generally do not support Euarchontoglires (e.g. Nikaido et al. 2001) or even maintain rodent polyphyly in many cases (Reyes et al. 2000; Arnason et al. 2002; Janke et al. 2002). Only by excluding some taxa with high or atypical substitution rates (or both) can sound mitochondrial support be obtained (Waddell et al. 2001). Establishing the monophyly of the most speciose eutherian order, Rodentia, and finding its sister group has indeed been most difficult to solve on the basis of sequence evidence (e.g. Graur et al. 1991; Adkins et al. 2001; Huchon et al. 2002). As for the molecular data sets giving support to Euarchontoglires, it has been questioned whether these are actually able to resolve the relationship of rodents and primates or whether more genes and longer sequences are needed (Rosenberg and Kumar 2001). Given, too, that Euarchontoglires is the least supported of the four major clades in some analyses (Madsen et al. 2001), additional evidence for their monophyly is certainly needed. This could be provided by ‘‘rare genomic changes,’’ such as insertions and deletions (indels) in proteins (Rokas & Holland 2000). Indels in

protein-coding DNA sequences require more complex mutational mechanisms and are generally more constrained than single base substitutions. Such indels can therefore be good indicators for monophyly, as demonstrated already for two of the other major clades, Xenarthra (van Dijk et al. 1999) and Afrotheria (Madsen et al. 2001), as well as in deeper vertebrate phylogeny (Venkatesh et al. 2001).

While studying genes involved in various neurodegenerative disorders, we noticed two deletions that might be informative for the naturalness of Euarchontoglires. One is a large deletion in exon 8 of the gene for spinocerebellar ataxia 1 (*SCA1*), resulting in an 18- residue deletion in the encoded protein (Fig.1, top). The other is a 6-bp deletion at the 59 end of the intronless coding region of the prion protein gene (*PRNP*; Fig.1, bottom). Both deletions perfectly distinguish Euarchontoglires from all other placentals and outgroup marsupials. Obviously, the most parsimonious interpretation is that these deletions originated once and independently in the *SCA1* and *PRNP* genes of the last common ancestor of Euarchontoglires, thus supporting their monophyly. If the morphological or mitogenomic trees are true, both deletions must have originated at least twice in exactly the same lineages.

Although reversal of the observed deletions in *SCA1* and *PRNP* is difficult to imagine, a repeated origin cannot totally be excluded. Indels are certainly not free from homoplasy, especially in regions with sequence repeats. In the *SCA1* gene, for example, a sequence repeat CTG TCN CCC, coding for Leu-Ser-Pro (underlined in Fig.1, top), might in principle have triggered the large deletion more than once. In the middle of this same region, a 6-bp deletion has caused the loss of two alanines in armadillo, whereas a 3-bp insertion results in an additional alanine in most Laurasiatheria (Fig.1, top). This latter insertion might indeed agree nicely with a basal separation of Eulipotyphla (represented here by hedgehog and mole) from the other Laurasiatheria (Murphy et al. 2001b). However, both the deletion and the insertion are likely to be caused by the GCC (Ala) repeat in this gene region and therefore to have little phylogenetic significance. It is the congruence of independent evidence that makes the two deletions as shown in figure 1 convincing indicators for the monophyly of Euarchontoglires. The probability of parallel origins of such deletions in two independent genes is difficult to evaluate statistically (van Dijk et al. 1999; Rokas & Holland 2000), but certainly it is extremely small. And even if these deletions were due to homoplasy, it would be a most curious coincidence that they occur in precisely

the same species that are also grouped by independent sequence evidence (Madsen et al. 2001; Murphy et al. 2001a, 2001b).

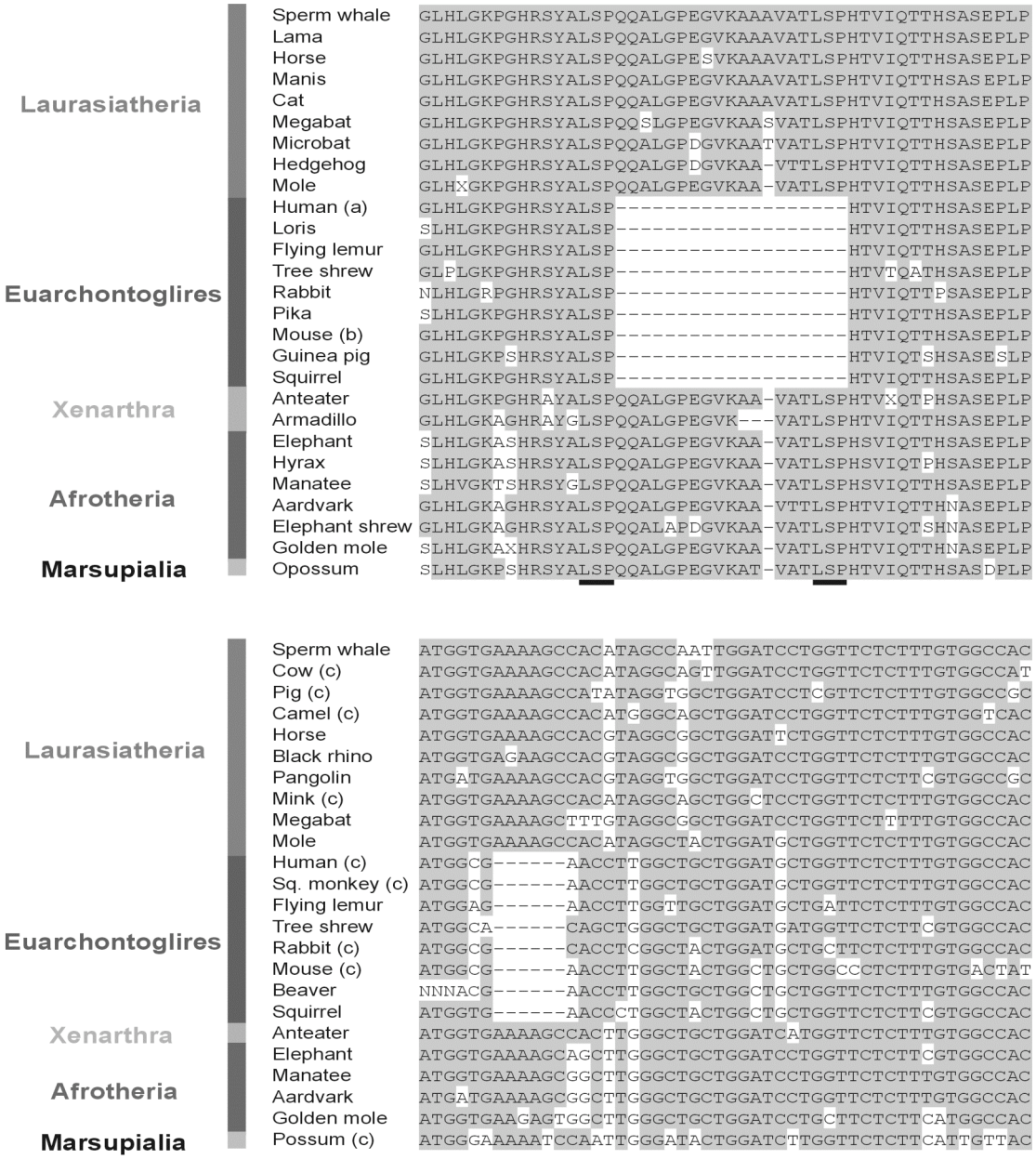

Figure 1: Deletions in the SCA1 protein (top) and the prion protein gene (bottom) support Euarchontoglires. Protein and DNA sequences, respectively, are shown as being most informative. Sequences correspond with positions 415 to 445 in the human SCA1 protein, and with nucleotides 1–44 of the coding sequence of the human *PRNP* gene. Eutherian species are grouped according to the four recently proposed basal clades of placental mammals (Murphy et al. 2000b). Gray shading emphasizes the overall sequence conservation; — denotes alignment gaps. The underlined Leu-Ser-Pro repeat in SCA1 is discussed in the text. Most sequences were newly determined by direct sequencing of PCRamplified genomic DNA fragments and can be found with full species names under accession numbers AJ438463–AJ438487 for *SCA1* and AJ438193–AJ438207 for *PRNP*. Human and mouse *SCA1* sequences are from the database (*a*, XM004164; *b*, NM009124), and *PRNP* sequences indicated with *c* from Wopfner et al. (1999).

ACKNOWLEDGMENTS

This work was supported by grants from the Netherlands Organization for Scientific Research and the European Commission.

REFERENCES

Adkins, R. M., E. L. Gelke, D. Rowe, and R. L. Honeycutt. (2001). Molecular phylogeny and divergence times for major rodent groups: evidence from multiple genes. *Mol. Biol. Evol.* 18: 777–791.

Arnason, U., J. A. Adegoke, K. Bodin, E. W. Born, Y. B. Esa, A. Gullberg, M. Nilsson, R. V. Short, X. XU, and A. Janke. (2002). Mammalian mitogenomic relationships and the root of the Eutherian tree. *Proc. Natl. Acad. Sci. USA* 99: 8151–8156.

Graur, D., W. A. Hide, and W.-H. Li. (1991). Is the guinea-pig a rodent? *Nature* 351:649–652.

Huchon, D., O. Madsen, M. J. J. Sibbald, K. Ament, M. J. Stanhope, F. Catzeflis, W. W. de Jong, and E. J. P. Douzery. (2002). Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* 19: 1053-1065.

Janke, A., O. Magnell, G. Wieczorek, M.Westerman, and U. Arnason. (2002). Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. *J. Mol. Evol*. 54: 71–80.

Liu, F.G., M M. Miyamoto, N P. Freire, P.Q. Ong, M.R. Tennant, T.S. Young and K.F. Gugel. (2001). Molecular and morphological supertrees for eutherian (placental) mammals. *Science* 291: 1786-1789.

Madsen, O., M. Scally, C.J. Douady, D.J. Kao, R.W. DeBry, R. Adkins, H.M. Amrine, M.J. Stanhope, W.W. de Jong and M.S. Springer. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409: 610-614.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder and S. J. O'Brien. (2001a). Molecular phylogenetics and the origins of placental mammals. *Nature* 409: 614-618.

Murphy, W. J., E. Eizirik, S. J. O'Brien et al. (2001b). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294: 2348-2351.

Nikaido, M., K. Kawai, Y. Cao, M. Harada, S. Tomita, N. Okada and M. Hasegawa. (2001). Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53: 508-516.

Novacek, M.J. (2001). Mammalian phylogeny: genes and supertrees. *Curr. Biol.* 11: R573-R575.

Reyes, A., G. Pesole and C. Saccone. (2000). Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. *Gene* 259: 177-187.

Rokas, A. and P.W. Holland. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15: 454-459.

Rosenberg, M.S. and S. Kumar. (2001). Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl. Acad. Sci. USA* 98: 10751-10756.

van Dijk, A. M., E. Paradis, F. Catzeflis and W. W. de Jong. (1999). The virtues of gaps: xenarthran (edentate) monophyly supported by a unique deletion in alpha-A-crystallin. *Syst. Biol*. 48: 94-106.

Venkatesh, B., M. V. Erdmann and S. Brenner. (2001). Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Natl. Acad. Sci. USA* 98: 11382-11387.

Waddell, P. J., H. Kishino, and R. Ota. (2001). A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics* 12: 141–154.

Wopfner, F., G. Weidenhofer, R. Schneider, A. von Brunn, S. Gilch, T. F. Schwarz, T. Werner and H. M. Schatzl. (1999). Analysis of 27 mammalian and 9 avian PrPs reveals high conservation of flexible regions of the prion protein. *J. Mol. Biol*. 289: 1163-1178.

Chapter 3

Primate Phylogeny, Evolutionary Rate Variations, and Divergence Times: A Contribution From the Nuclear Gene IRBP

Céline Poux[1,2] and Emmanuel Douzery[2]

[1]Laboratoire de Paléontologie, Paléobiologie et Phylogénie-CC064, Institut des Sciences de l'Evolution UMR 5554/ CNRS, Université Montpellier II 34095 Montpellier, France

[2]Department of Biochemistry, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

American Journal of Physical Anthropology 124:1–16. 2004

ABSTRACT

The first third (ca. 1200 bp) of exon 1 of the nuclear gene encoding the Interstitial Retinoid Binding Protein (IRBP) has been sequenced for 12 representative primates belonging to Lemuriformes, Lorisiformes, Tarsiiformes, Platyrrhini and Catarrhini, and combined with available data (13 other Primates, 11 non-primate placentals, and 2 marsupials). Phylogenetic analyses using maximum likelihood on nucleotides and amino acids robustly support the monophyly of Primates, Strepsirrhini, Lemuriformes, Lorisiformes, Anthropoidea, Catarrhini, and Platyrrhini. It is interesting to note that 1) Tarsiidae grouped with Anthropoidea, and the support for this node depends on the molecular characters considered; 2) Cheirogaleidae grouped within Lemuriformes; and 3) *Daubentonia* was the sister-group of all other Lemuriformes. Study of the IRBP evolutionary rate shows a high heterogeneity within placentals and also within Primates. Maximum likelihood local molecular clocks were assigned to three clades displaying significantly contrasted evolutionary rates. Paenungulata were shown to evolve 2.5 to 3 times faster than Perissodactyla and Lemuriformes. Six independent calibration points were used to estimate splitting ages of the main primate clades, and their compatibility was evaluated. Divergence ages were obtained for the following crown groups: 13.8-14.2 MY for Lorisiformes, 26.5-27.2 MY for Lemuroidea, 39.6-40.7 MY for Lemuriformes, 45.4-46.7 MY for Strepsirrhini, and 56.7-58.4 MY for Haplorrhini. The incompatibility between some paleontological and molecular estimates may either reflect the incompleteness of the placental fossil record, and / or indicate that the variable IRBP evolutionary rates are not fully accommodated by local molecular clocks.

INTRODUCTION

Primates form one of the 18 orders of placental mammals identified by morphology, paleontology (Novacek 1992a) and molecular data (de Jong 1998). The extant sister group of primates is not clearly identified. On morphological bases like the ankle structure (Novacek 1992b, 1994), primates are grouped with Chiroptera (microbats and megabats), Scandentia (tree shrews) and Dermoptera (flying lemurs) into Archonta. Molecular studies suggest the polyphyly of this superorder due to the

inclusion of Chiroptera (Adkins & Honeycutt 1991; Stanhope et al. 1992; Bailey et al. 1992; Ammerman & Hillis 1992; Allard et al. 1996; Murphy et al. 2001a). After exclusion of Chiroptera, the phylogenetic relationships between the three remaining suborders (Euarchonta, Waddell et al. 1999) are still debated. Furthermore, in recent molecular studies, Primates are either linked with Glires (Madsen et al. 2001), or are not monophyletic (Murphy et al. 2001a; Arnason et al. 2002) due to the inclusion of Dermoptera (but for contrasting results, see Eizirik et al. 2001). The newest survey (Murphy et al. 2001b) displays strong support for Primates as sister group of Dermoptera plus Scandentia. This question might be clarified by an extensive taxon sampling of both Primates and non-primate placentals.

Despite the abundance of studies on primates, some questions about their phylogeny remain unanswered. The most debated point is about the evolutionary position of Tarsiiformes. Indeed, Primates are divided into two suborders of which the taxon content differs according to the phylogenetic position of tarsiers. *Tarsius* shares morphological characters with lemuriforms, lorisiforms, and anthropoids. Consequently, *Tarsius* is either grouped with Malagasy lemuriforms and Afro-Asian lorisiforms into prosimians, while the remaining primates form the simians (Rowe 1996; Murphy et al. 2001a,b), or *Tarsius* is associated with anthropoids (represented by Afro-Asian catarrhines and South-American platyrrhines) to form the haplorrhines (Goodman et al. 1998; Zietkiewicz et al. 1999; Schmitz et al. 2001), while other primates (lemuriforms and lorisiforms) are classified into strepsirrhines. The latter group presents many morpho-anatomical synapomorphies (the dental toothcomb and the laterally flaring talus) and symplesiomorphies (the moist rhinarium, the tapetum, and the bicornuate uterus) (Fleagle 1999).

Within strepsirrhines, both lorisiforms and lemuriforms appear monophyletic (Yoder et al. 1996a, b; Yoder 1997; Goodman et al. 1998). In the lorisiform clade, Galagonidae is monophyletic whereas the status of Loridae is controversial between morphological and molecular studies (Yoder et al. 2001). In the lemuriform clade, each family (Lemuridae, Cheirogaleidae, Daubentoniidae, Indridae, Lepilemuridae) is well-defined; nevertheless the relationships among them can be represented as a multifurcation, with different studies suggesting different phylogenetic relationships (for a review of molecular studies, see Yoder 1997). The phylogenetic status of two families has been highly controversial. First, dwarf and mouse lemurs (Cheirogaleidae) are endemic to Madagascar, and considered as lemuriforms, but the

anatomy of their ascending pharyngeal artery would suggest associating them with lorisiforms (Szalay & Delson, 1979; but see Yoder 1994). Second, the aye-aye (Daubentoniidae: *Daubentonia*) is considered as the sister-group of either all other lemuriformes (Yoder et al. 1996a, 2003) or strepsirrhines on the basis of morphology (Groves 1989) and mitochondrial molecules (Adkins & Honeycutt 1994; Arnason et al. 1998). The latter hypothesis would all involve a double event of colonization of Madagascar by lemuriforms, a biogeographic scenario that is refuted by other molecular studies (Yoder 1994, 1997; Yoder et al. 1996a, 2003).

About platyrrhines (the South American anthropoids), molecular phylogenies are not in agreement with morphological classifications. Platyrrhines (i.e. ceboids) are traditionally divided into the two families Callitrichidae (marmosets and tamarins) and Cebidae (cebids) (Rowe 1996). Molecular studies tend, however, to demonstrate that cebids are paraphyletic, because the Cebinae subfamily would group with callitrichids (Goodman et al. 1998). The relationships among Pithecidae (e.g. *Pithecia*), Atelidae (e.g. *Ateles*), and Cebidae (Cebinae, e.g. *Cebus* or *Saimiri*, and Callitrichinae, e.g. *Callithrix*) remain unresolved (Goodman et al. 1998; Canavez et al. 1999; von Dornum & Ruvolo 1999; Schneider 2000; Schneider et al. 2001). A composite primate phylogeny reconstructed from many independent studies also found different results (Purvis 1995).

The timing of the evolution of primates is also debated. For example, mitochondrial markers used to assess primate divergence dates display ages far more ancient than the paleontological ones (Arnason et al. 1998, 2000; Yoder & Yang, 2000; vs. Gingerich 1984; Gingerich & Uhen 1994, Fleagle 1999, Rosenberger et al. 1991). Gaps in the fossil record of Primates and variation in molecular evolutionary rates between lineages, e.g. faster rates in anthropoid primates (Adkins & Honeycutt 1994; Andrews et al. 1998; Andrews & Easteal 2000; Liu et al. 2001) and slower rates in strepsirrhines (Yoder et al. 1996a), might explain this observation. Because the absence of a global molecular clock still remains one of the most limiting factors of molecular dating, two different approaches were recently proposed and applied to estimate primate speciation ages. Local molecular clocks in a maximum likelihood framework allocate independent substitution rates to groups that evolve at different rates (Yoder & Yang 2000), and relaxation of the molecular clock is achieved in a Bayesian framework to describe evolutionary rate variation along tree nodes (Thorne et al. 1998; Yoder et al. 2003).

To reconstruct the phylogeny of primates and to estimate their divergence times, we focused on a nuclear marker because of its potential resolving power relative to mitochondrial genes at deep (Springer et al. 2001) and ordinal (DeBry & Sagel 2001) levels. The nuclear gene for the Interstitial Retinoid-Binding Protein (IRBP) encodes a 140-kD protein involved in vision. It passively carries the retinol from the photoreceptors to the pigmented epithelium, where it is transformed in retinal (references *in* Nickerson et al. 1998), but its exact function is not yet well established. The IRBP gene is exclusively present in the genome of vertebrates (Borst et al. 1989). It is located on human chromosome 10 (Fong et al. 1990) and comprises four exons. In mammals, the first exon is formed by three repeated regions (each corresponding to ca. 300 amino acids) and the beginning of a fourth region (Wagenhorst et al. 1995). The first 1.2 kb of exon 1, which covers the first and half of the second repeat, were sequenced for a variety of placental and marsupial mammals, and are widely used to reconstruct their phylogeny (e.g. Stanhope et al. 1992, 1996; Springer et al. 1997; Jansa & Voss 2000; DeBry & Sagel 2001; Madsen et al. 2001; Huchon et al. 2002; Mercer & Roth 2003). IRBP displays several advantages: 1) it is a single copy nuclear gene, and no other gene belonging to the same potential family has been identified (Borst et al. 1989); 2) the high degree of divergence between the repeats (up to 60%) avoids confusion between them, and allows us to compare orthologous sequences; 3) the coding nature of the IRBP sequence allows to compare taxa at different taxon levels, e.g. within mammalian orders (Jansa & Voss 2000; DeBry & Sagel 2001), including primates (Yoder & Irwin 1999; Yoder et al. 2001, 2003); and 4) a large dataset of mammalian sequences is available. Here, we use the nuclear gene for IRBP to study the phylogenetic relationships within all major clades of primates. We provide evidence of marked evolutionary rate differences for this gene in primates, and we apply a maximum likelihood approach based on local molecular clocks to estimate the divergence times of their major clades.

MATERIAL & METHODS

Taxon sampling

We incorporated 25 primate species in our analyses, with at least one representative for each family, except Pongidae (Table 1). We added two marsupials (*Didelphis virginiana:* Virginia opossum, and *Macropus giganteus:* eastern gray kangaroo), and 11 nonprimate placentals sampled from the four placental lineages identified by Murphy et al. (2001a): 1) Afrotheria, represented by *Dugong dugon* (dugong), *Elephantulus rufescens* (rufous elephant shrew), *Loxodonta africana* (African elephant), and *Procavia capensis* (rock hyrax); 2) Xenarthra: *Bradypus tridactylus* (pale-throated three-toed sloth); 3) Euarchontoglires (Euarchonta + Glires), including potential sister groups of primates: Dermoptera (*Cynocephalus variegatus:* Malayan flying lemur), Scandentia (*Tupaia glis:* common tree shrew), Lagomorpha (*Oryctolagus cuniculus:* European rabbit), and Rodentia: (*Mus musculus:* house mouse); and 4) Laurasiatheria, represented by Perissodactyla: *Equus caballus* (horse) and *Tapirus pinchaque* (mountain tapir). This taxonomic sampling is justified by the fact that primates belong to the unambiguously monophyletic Euarchontoglires (Madsen et al. 2001; Murphy et al. 2001a,b; Poux et al. 2002). Laurasiatherians, the xenarthran, afrotherians, and the marsupials are included as successively more distant outgroups, and the choice of the laurasiatherian and afrotherian representatives is dictated by their ability to provide paleontological calibration points for molecular dating (see below).

DNA amplification and sequencing of exon 1 of IRBP

DNA of 12 primate species (Table 1) was extracted from tissue samples in the collection of 95% ethanol preserved mammalian tissues of the Institut des Sciences de l'Evolution (Catzeflis 1991). Nucleotide sequences were obtained for the partial exon 1 of the IRBP gene (1278 bp, corresponding to positions 25–451 of the protein) by polymerase chain reaction (PCR). Two fragments with 300 overlapping base pairs (bp) were amplified: I1/J2 (827 bp) and I2/J1 (931 bp), using primers I1 (5'-ATGGCCAAGGTCCTCTTGGATAACTACTGCTT-3'), J1 (5'-CCACTGCCCTCC-

CATGTCTG-3'), I2 (5'-ATCCCCTATGTCATCTCCTACYTG- 3'), and J2 (5'-CGCAGGTCCATGATGAGGTGCTCCGTGTCCTG-3').

Table 1. **Latin and common names, tissue collectors, and IRBP EMBL-GenBank-DDBJ accession numbers for the primate taxa used in this study.**

Latin name	Common name	Collector or Reference	Accession
ANTHROPOIDEA			
CATARRHINI			
Hominoidea			
Hylobates lar[*]	White-handed Gibbon	Ulfur Arnason	AJ313478
Homo sapiens	Human	(Fong et al., 1990)	J05253
Cercopithecoidea			
Cercopithecus solatus[*]	Sun-tailed Guenon	Jean-Pierre Hugot	AJ313477
Macaca mulatta[*]	Rhesus Macaque	Dr. M. Brack ; Deutsches Primaten Zentrum	AJ313476
PLATYRRHINI			
Ateles paniscus[*]	Black Spider Monkey	Jean-François Mauffrey	AJ313474
Callithrix jacchus[*]	Common Marmoset	Dr. M. Brack ; Deutsches Primaten Zentrum	AJ313472
Cebus apella[*]	Brown Capuchin	Jean-François Mauffrey	AJ313473
Pithecia pithecia[*]	White-faced Saki	Faune Sauvage (EDF-CNEH)	AJ313475
Saimiri sciureus	Common Squirrel Monkey	(Yoder et al., 2001)	AF271424
TARSIIFORMES			
Tarsius bancanus	Western Tarsier	(Yoder et al., 2001)	AF271423
Tarsius syrichta	Philippine Tarsier	(Stanhope et al., 1992)	Z11806
STREPSIRRHINI			
LEMURIFORMS			
Lemuroidea			
Lemur catta[*]	Ring-tailed Lemur	Mr. Combes Zoo Montpellier	AJ313470
		(Yoder and Irwin, 1999)	AF081058
Hapalemur griseus	Lesser Bamboo Lemur	(Yoder and Irwin, 1999)	AF081057
Varecia variegata	Ruffed Lemur	(Yoder and Irwin, 1999)	AF081056
Eulemur mongoz	Mongoose Lemur	(Yoder and Irwin, 1999)	AF081064
Microcebus murinus[*]	Gray Mouse Lemur	Noëlle Bons	AJ313469
		(Yoder and Irwin, 1999)	AF081054
Cheirogaleus major	Greater Dwarf Lemur	(Yoder et al., 2001)	AF271421
Propithecus verreauxi[*]	Verreaux's Sifaka	R. Albignac	AJ313471
Propithecus tattersalli	Tattersall's Sifaka	(Yoder and Irwin, 1999)	AF081053
Daubentonioidea			
Daubentonia madagascariensis[*]	Aye-aye	Michel Tranier, MNHN; (Yoder et al., 2001)	AJ313468 AF271422
LORISIFORMS			
Nycticebus coucang[*]	Slow Loris	Ole Madsen	AJ313467
		(Yoder et al., 2001)	AF271419
Loris tardigradus	Slender Loris	(Yoder et al., 2001)	AF271418
Perodicticus potto	Potto	(Yoder et al., 2001)	AF271420
Otolemur crassicaudatus	Thick-tailed Greater Bush Baby	(Stanhope et al., 1992)	Z11805
Galagoides demidoff	Demidoff's Bush Baby	(Yoder et al., 2001)	AF271416
Galago moholi	Southern Lesser Bush Baby	(Yoder et al., 2001)	AF271415

[*] Taxa sequenced in the present study.

PCR reactions were performed using the followings parameters: 29 cycles with 94°C denaturation (20 sec), 47°C annealing (30 sec), 68°C extension (2 min), and one final cycle of 68°C extension (10 min). A minimum of two PCR products was pooled and excised from a 1% agarose gel in TAE 1x buffer, and then purified on Ultrafree-DA Amicon columns (Millipore), and reconcentrated on Microcon filterable columns (Millipore). Manual sequencing was conducted using the dideoxy chain termination method with [α^{33}P-ddNTP] and the Thermo Sequenase cycle sequencing kit (Amersham). PCR fragments were sequenced on both strands with I1/J2 and I2/J1 external primers and with internal primers I5 (5'-GCCCTGGACCTCCAGAAGCT-GAGGATMGG-3') and J5 (5'-CARGGTCCAGATCTCYGTGGT-3').

Phylogenetic analyses

IRBP sequences were aligned by hand with the ED editor of the MUST package (Philippe 1993), version 2000. Sites not sequenced for more than 75% of the taxa were removed from subsequent analyses. Other nonsequenced positions and gap sites were coded as missing data. Phylogenetic reconstructions were performed by maximum likelihood (ML) with PAUP* (Swofford 2001), version 4, beta 8. ML was exclusively used because it is a powerful technique based on explicit models of sequence evolution that allows statistical testing of alternative phylogenetic hypotheses (Whelan et al. 2001).

The ML assumptions included a general time reversible (GTR) model of nucleotide sequence evolution, and an eight-category Gamma distribution (Γ_8) to describe the substitution rate heterogeneities between sites (Yang 1996a). Maximum likelihood parameters of the GTR+ Γ_8 model were estimated by PAUP*, and the highest-likelihood topology was identified after an ML heuristic search conducted with a neighbor-joining (NJ) starting tree, and tree bisection-reconnection (TBR) branch swapping. The stability of the nodes was estimated by bootstrap (Felsenstein 1985), with 500 replicates of heuristic searches (NJ starting trees, ML parameters identically set to their optimal value for each replicate, and TBR branch swapping with limitation to 1,000 rearrangements per replicate). ML analysis of amino acids was conducted with PAML (Yang 1997), version 3.0d, using nearest-neighbor interchange (NNI) branch swapping.

The choice of the model of DNA evolution was justified because the log-likelihood of the best ML tree estimated by PAUP* increases from lnL = -13,383.70 in the HKY model, to -13,383.61 in the TN93 model (δ = 0.09; P = 0.67 for the significance of the more complex model under the likelihood ratio test), -13,372.74 in the GTR model (δ = 10.87; $P < 0.001$), -12,610.55 in the GTR + Γ_8 model (δ = 762.19; $P < 0.001$), and -12,608.53 in the GTR + Γ_8 model with a fraction of invariable sites (δ = 2.02; P = 0.05). To homogenize analyses between PAUP* and PAML, we used the GTR + Γ_8 model for all DNA analyses, the fraction of invariable sites being not implemented under PAML. For protein evolution, we used the Jones-Taylor-Thornton (JTT) + Γ_8 model, with amino-acid frequencies adjusted to the IRBP data set (+ F option).

Test of alternative phylogenetic hypotheses

Alternative hypotheses were evaluated in an ML framework, using the nonparametric KH test (Kishino & Hasegawa 1989), with correction for comparisons of topologies defined a posteriori (KH-SH test) (Shimodaira & Hasegawa 1999). To take into account the different evolutionary processes of each codon position (in terms of nucleotide frequencies, substitution rates, and rate variation across sites; cf. Table 2), we followed the approach suggested by Yang (1996b). A partitioned likelihood analysis was conducted, with GTR + Γ_8 model parameters independently estimated for each of the three IRBP codon positions and independence of branch lengths estimated across the three codon positions. To appreciate the phylogenetic content of third codon positions (Yoder et al. 1996b; Yoder & Yang 2000) and their impact on the acceptance or rejection of evolutionary alternatives, KH-SH tests were performed: 1) on all codon positions, 2) on nucleotides after exclusion of third codon positions, and 3) on amino acids. All tests were performed with PAML 3.0d., with estimation of three sets of base composition, GTR rates, Gamma shape, and branch lengths parameters, one for each codon position.

Alternative phylogenetic hypotheses for *Tarsius* were also evaluated under the parametric SOWH test (Swofford et al. 1996), following the guidelines for the "posPfud" procedure of Goldman et al. (2000): 1) The alternative topology to be tested was a posteriori defined, e.g. *Tarsius* was constrained to be sister group of the

Strepsirrhini (the Prosimia topology). 2) The parametric approach was conducted by simulating 1,000 character matrices under Seq-Gen (Rambaut & Grassly 1997), version 1.2.5, using the Prosimia topology and its optimal likelihood GTR, Γ_8, and branch length parameters. 3) The loglikelihoods of three topologies (Prosimia (T_P), *Tarsius* in basalmost position (T_B) among primates, and Haplorrhini (T_H)) were computed under the 1,000 simulated matrices, with a full optimization method where ML parameters were estimated from the data. 4) The SOWH test was conducted by comparison against the uncentered distribution of the 1,000 parametric estimates of the difference in log-likelihoods of the best tree (T_P, T_B, or T_H) and that of the alternative tree (T_P). 5) The confidence level of the test was obtained by direct comparison of the test statistics with the estimated distribution.

Table 2. Molecular characteristics of the IRBP exon 1 inferred from partitioned maximum likelihood analysis. The following parameters are given for each codon position and for their combination: total, variable, and informative number of characters; base composition in percent of A, C, G, and T; the relative substitution rate of each partition calculated relative to the slowest; rate parameters of the GTR model of sequence evolution; and the a parameter of the gamma distribution of the rate heterogeneity among sites.

	IRBP codon positions			
	First	**Second**	**Third**	**All**
Total number of characters	421	421	421	1263
Variable characters	233	159	389	781
Informative characters	152	99	338	589
%A	20.4	26.1	9.8	18.8
%C	29.4	24.5	40.3	31.4
%G	38.8	18.6	38.4	32.0
%T	11.3	30.8	11.4	17.8
Relative rate	1.49	1.00	5.62	—
A ↔ C	1.25	3.20	1.37	1.56
A ↔ G	2.97	13.22	8.18	6.30
A ↔ T	0.87	0.94	4.73	1.35
C ↔ G	0.58	3.19	0.49	1.04
C ↔ T	3.36	4.69	8.38	6.27
G ↔ T	1.00	1.00	1.00	1.00
a (G distribution)	0.72	0.37	2.98	0.53

Local clock analyses

A three-step approach was conducted for molecular dating on DNA characters. First, we detected species or clades that evolved significantly slower or faster than the others by three complementary approaches. The two-cluster (TC) and branch-length (BL) tests from the LINTRE package (Takezaki et al. 1995) respectively examined 1) the hypothesis of equality of the average substitution rate

for two clusters that are linked by a given node in the tree, and 2) the deviation from the average of the total branch lengths connecting the root of the ingroup to a given terminal sequence. Relative-rate tests between groups of sequences were conducted with RRTree (Robinson-Rechavi & Huchon 2000), to take into account the phylogenetic relationships between taxa and to investigate substitution rate differences between clades that are not directly connected in the phylogeny (Robinson et al. 1998). We evaluated rate variations between six placental clades (primates, *Cynocephalus* + *Tupaia,* Glires, Perissodactyla, Paenungulata, and Xenarthra). Rate variations were also explored within primates and involved five clades (Lemuriformes, Lorisiformes, Tarsiiformes, Platyrrhini, and Catarrhini). Dermoptera and Scandentia were chosen as the nearest outgroup which increases the power and accuracy of the relative rate test (Robinson et al. 1998).

Second, we followed the ML local molecular clock approach of Yoder & Yang (2000), because coding regions of the nuclear genome, like IRBP, are often subject to important variations in evolutionary rates among primates (Bailey et al. 1991; Liu et al. 2001). This approach postulates different evolutionary rates for some lineages while assuming (local) rate constancy in others. It represents a compromise between two extreme situations, that either use a global molecular clock (a single substitution rate is imposed for all lineages), or independent rates for each branch (no clock is imposed). In this latter case, molecular dating will be highly sensitive to potential evolutionary rate differences between lineages. Different evolutionary rates were assigned under ML to taxa and clades previously identified to be deviating by the relative rate tests, in order to obtain a tree that satisfies the local clock hypothesis. Likelihood ratio tests (Felsenstein 1988) were performed to test whether the IRBP nucleotide sequences fit global or local molecular clock hypotheses. One should note that the "ML local clock" approach of Yoder & Yang (2000) substantially differs from the "local clock" approach of Bailey et al. (1991): the former is based on the likelihood criterion to define local constancy of rates, whereas the latter is based on a distance approach and reiterates a calibration of each new local clock, based on the molecular time estimate of the previous (deeper) node.

Third, we used six independent divergence points consistent with the paleontological record in order to calibrate the local molecular clocks. Among the factors influencing the reliability of molecular dating, it was found that one of the most important is the choice of calibration points (Huchon et al. 2000; Yoder & Yang,

2000). Therefore, cross-calibration comparisons should be performed to evaluate the reliability of each calibration point (Huchon et al. 2000). Two calibration points were outside the primates: the radiation of paenungulates (C_{PAE}; see Fig.3) at 55–60 MY (Gheerbrant et al. 1996), and that of perissodactyls (C_{PER}) at 55 MY (Garland et al. 1993). Four points were chosen within the primates: the radiation of the order (C_{PRI}) at 63 MY (Gingerich 1984; Gingerich & Uhen 1994), of Anthropoidea (C_{ANT}) at 34 MY (Fleagle 1999), of Plathyrrhini (C_{PLA}) at 26MY (Rosenberger et al. 1991), and of Catarrhini (C_{CAT}) at 20–25 MY (Fleagle 1999). All local clocks, divergence dates, and standard errors were estimated with PAML (Yang 1997), version 3.0d.

RESULTS AND DISCUSSION

Molecular properties of IRBP

The nucleotide composition for the three IRBP codon positions was homogeneous for all taxa except *Macropus* ($P < 0.01$), *Didelphis* ($P < 0.01$), and *Elephantulus* ($P = 0.02$), as estimated by a χ^2 test of deviation from the mean. For this reason, all codon positions were kept in subsequent phylogenetic analyses. The partitioned maximum likelihood analysis yields information about the molecular characteristics of each codon position of IRBP, as presented in Table 2. Third-codon positions are evolving 5.62 faster than second positions. They provide most of the variable sites of the complete alignment, and substitutions occur at nearly all of the third positions, as indicated by the high value of the Γ-distribution parameter. Third positions also display a strong base composition bias against A and T, making this IRBP exon rather G + C-rich (78.7%). Transitions occur more frequently than transversions at all codon positions, and a highly pronounced bias towards A ↔ G transitions is recorded at second positions, as well as a rather high level of A ↔ C and C ↔ G transversions.

Monophyly of primates and their phylogenetic position within placentals

Because of ML computing time restrictions, the taxonomic sampling was limited to 36 placentals rooted with two marsupials. This included 12 new primate IRBP sequences, and a selection of laurasiatherians and afrotherians restricted to taxa providing calibration points for molecular dating. The ML analysis of all codon positions of this data set highly supports the monophyly of primates (bootstrap percentage (BP) = 92), here represented by an extended taxon sampling and including a variety of other placental representatives (Fig.1). The IRBP tree also depicts the major placental clades as previously identified by Murphy et al. (2001a,b) and Madsen et al. (2001): Afrotheria (BP = 100), *Bradypus* (representing Xenarthra), Perissodactyla (BP = 100, representing Laurasiatheria, and providing one fossil calibration point; see below), and a clade (BP = 43) containing primates, Dermoptera, Scandentia, and Glires (Rodentia + Lagomorpha). *Equus* plus *Tapirus* branch off at a basal position relative to other placentals, a feature consistent with the observations of DeBry & Sagel (2001) using the same molecule. However, this branching pattern receives no support (BP = 20), and is not in agreement with molecular studies in which Laurasiatheria are connected to Euarchontoglires (Madsen et al. 2001; Murphy et al. 2001a,b). This might be explained by a spurious rooting by the two divergent marsupial sequences. A KH-SH test performed to evaluate both hypotheses displayed no significant difference ($P = 0.23$ for nucleotide sequences). This is why we decided to constrain perissodactyls to branch as the sister clade of Euarchontoglires to calculate divergence dates. The flying lemur (*Cynocephalus*) and the tree shrew (*Tupaia*) cluster together (BP = 63), and are closer to Glires than to primates (BP = 49). This weakly supported grouping is consistent with the phylogeny of DeBry & Sagel (2001) using the same marker, but implies the paraphyly of Euarchonta (primates, Dermoptera, and Scandentia), a group usually considered monophyletic (Madsen et al. 2001; Murphy et al. 2001a,b). IRBP unambiguously suggests the monophyly of primates, but does not answer the question of their sister clade, perhaps because of the limited number of molecular characters and/or poor taxon sampling within Glires.

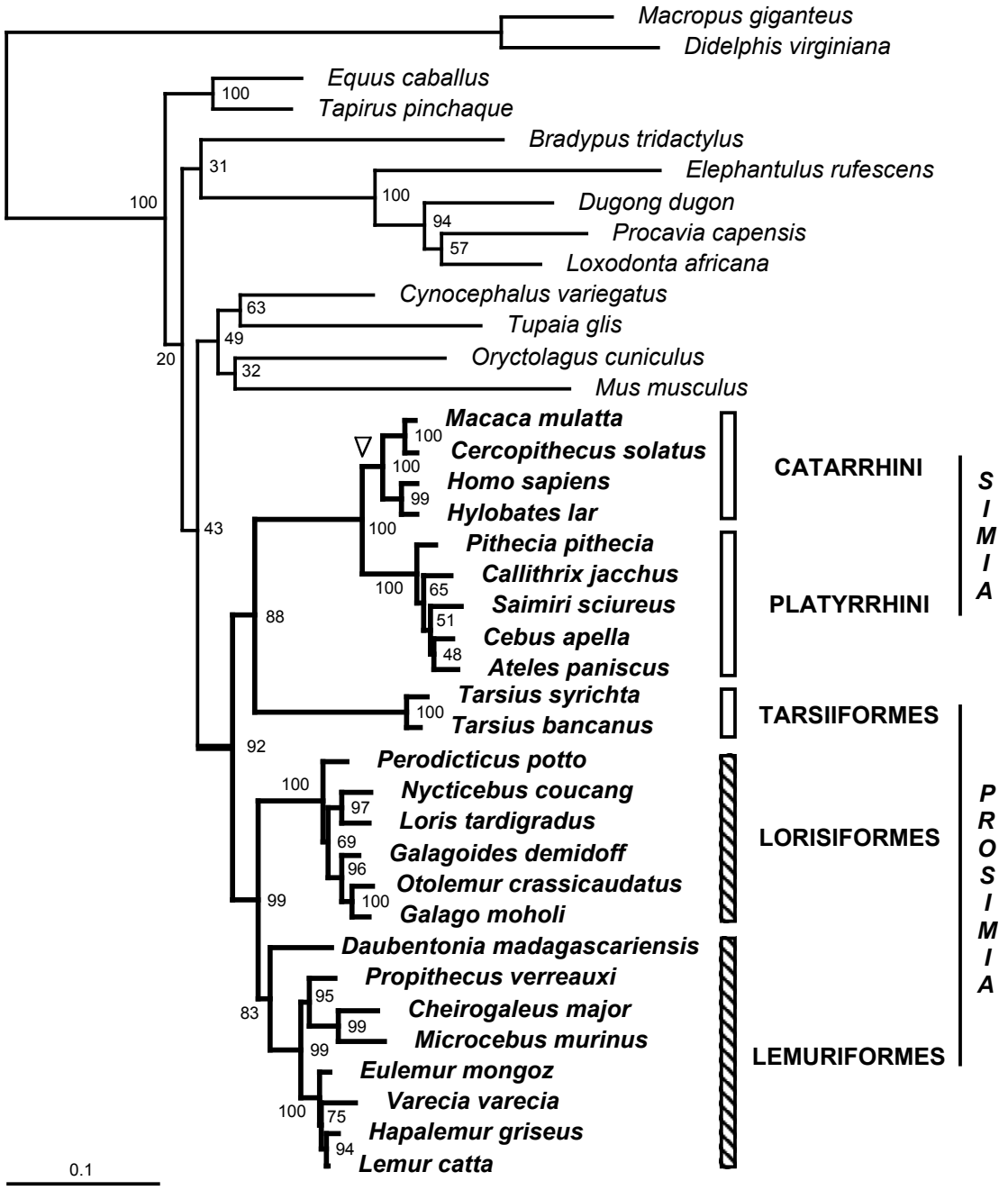

Figure 1: Highest-likelihood phylogram (-ln*L* = 12,297.21) of primate and other placental relationships reconstructed from all codon positions of nuclear gene IRBP. Branch lengths are proportional to number of predicted substitutions per site, under a GTR model with rate matrix (AC; 1.58; AG; 6.27; AT; 1.31; CG; 1.03; CT; 6.23; GT; 1.00), and a Gamma distribution of parameter α = 0.53. Branch leading to marsupial outgroup was reduced four times in length. Maximum likelihood bootstrap percentages obtained after 500 replicates are indicated at nodes. Open triangle indicates that one insertion of two consecutive codons is diagnostic for monophyly of catarrhines. Open and hatched rectangles indicate Haplorrhini and Strepsirrhini taxa, respectively. Simia is monophyletic, whereas Prosimia is paraphyletic due to position of Tarsiiformes.

Phylogenetic relationships within primates

IRBP polymorphism between closely related taxa

Yoder and Irwin (1999) and Yoder et al. (2001) sequenced a shorter (939 bp) but overlapping part of IRBP for a number of individuals belonging to the same or congeneric species, as sequenced in the present study. This allows one to evaluate the IRBP polymorphism between some pairs of identical and closely related taxa over 939 common positions: *Lemur catta* (0.3% nucleotide divergence), *Microcebus murinus* (2.0%), *Daubentonia madagascariensis* (0.6%), *Nycticebus coucang* (0.6%), and *Propithecus verreauxi*/*P. tattersalli* (0.7%). Some ambiguities, likely due to sequencing errors, in the last 130 bp of the *Microcebus* IRBP sequence of Yoder and Irwin (1999) might explain the high level of intraspecies polymorphism observed in this species. Actually, the removal of this DNA region leads to 0.6% divergence, which conforms to the values (less than 1.0%) observed for the other taxa. This higher nucleotidic divergence could be also explained by a phylogeographic variation in *Microcebus.* Indeed, phylogenetic analysis of mtDNA sequences showed a high diversity in this species (Yoder et al. 2000). Overall, the low IRBP exon 1 polymorphism observed between closely related taxa suggests that the phylogenetic relationships inferred between primate genera (see below) have not been affected by this factor.

Phylogenetic position of Tarsius

The ML analysis of the IRBP sequences recovered several major clades and subclades of primates with high support (Fig.1): Strepsirhini (BP = 99), Lorisiformes (BP = 100), Lemuriformes (BP = 83), *Tarsius* = Anthropoidea (BP = 88), Anthropoidea (BP = 100), Catarrhini (BP = 100), and Platyrrhini (BP = 100). Interestingly, the IRBP gene suggests the monophyly of Haplorrhini (Tarsiiformes + Anthropoidea).

Whatever the characters used (nucleotides with or without third-codon positions), the two alternative topologies for the branching position of *Tarsius* (i.e. Prosimia and *Tarsius* basal among primates) all involve a decrease in log-likelihood, which is highly significant ($P < 0.001$) under the parametric SOWH test (Fig.2).

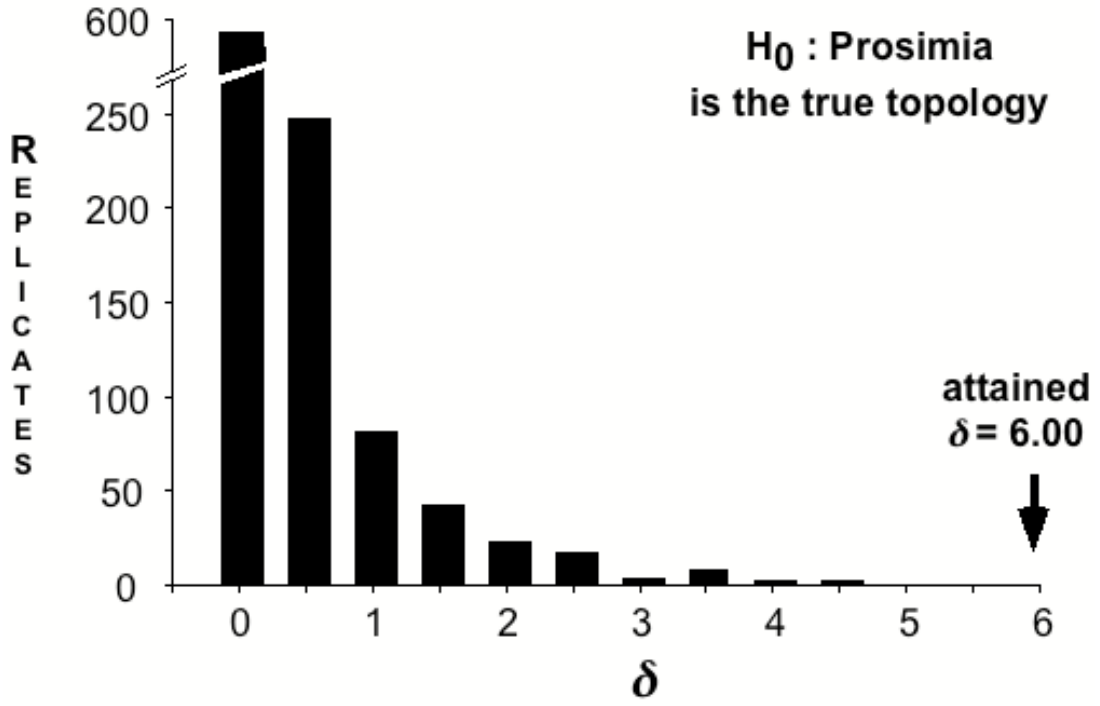

Figure 2: Statistical rejection of Prosimia hypothesis by SOWH test conducted on all codon positions of IRBP. After evaluating 1,000 simulated data sets, greatest difference of log-likelihood between Prosimia and either Haplorrhini or basal *Tarsius* hypotheses was 4.07, which is lower than δ = 6.00 observed on original data set. Prosimia hypothesis is therefore rejected at $P < 0.001$.

Contrasting with these results, the log-likelihood drop for alternatives to the monophyly of Haplorrhini does not reach the significance level of 5% under the nonparametric KH-SH test, neither at the nucleotide nor at the amino-acid level (Table 3). As compared to Haplorrhini, the Prosimia hypothesis is always worse than the unorthodox grouping of *Tarsius* as sister group of all other Primates (i.e. Strepsirrhini and Anthropoidea are sister group), and this result is most pronounced when all nucleotide positions are evaluated (e.g. $P = 0.08$ vs. 0.18). The difference of behavior between the SOWH and KH-SH tests was previously noted, with the latter being more conservative. This might be explained by the increased power and/or greater reliance on sequence evolution models of parametric tests (Goldman et al. 2000).

The KH-SH tests of alternative phylogenetic hypotheses were conducted under partitioned ML. To compare models with a single character partition or with partitioning of the data according to codon positions, we used the Akaike information criterion (AIC), which should be minimized for the most suitable model. When a single model is defined for IRBP, the best tree (Fig.1) had a log-likelihood of lnL = –12,647.16 (as estimated by PAML), with a total of 82 free parameters (i.e. three

independent nucleotide frequencies, five independent GTR rates, one Gamma rate, and 73 [2 × 38 – 3] branches), and this gives an AIC of 2 × 12,647.16 + 2 × 82 = 25,458.32. When IRBP is partitioned according to codon positions, the highest log-likelihood is –12,295.94, with 3 × 82 = 246 free parameters, and this yields a lower AIC of 2 × 12,295.94 + 2 × 246 = 25,083.88. Therefore, the three-partition model better describes the IRBP data, and the decrease of log-likelihood observed when the monophyly of Tarsiiformes + Anthropoids is disrupted does not reflect the use of an oversimplified ML model (Whelan et al. 2001).

A critical analysis of IRBP phylograms after ML reconstruction of character-state changes shows that there are 13 synapomorphies supporting the Haplorrhini clade (five on first codon positions (P1), two on second (P2) positions, and six on third (P3) positions), against five (two on P1, two on P2, and one on P1) favoring the basal position of *Tarsius* among primates, and one supporting prosimians on P1. The latter one occurs on a site that, according to the polarization of the character-state change, can support the three alternative hypotheses. This means that this site cannot be used to discriminate between the three alternative branching positions of Tarsiiformes, and this implies that the IRBP gene does not support the monophyly of Prosimia.

The majority of shared derived substitutions for Haplorrhini thus occurred at third positions (most of them being silent), which is likely to explain that both removal of third positions and analysis of amino acids decrease the nonparametric statistical contrast between the three alternative topologies (see Table 3). These results (KH-SH tests, and the distribution of synapomorphies among different codon positions) show that third-codon positions might contain phylogenetic information, as previously suggested (e.g. Yoder et al. 1996b; Yoder & Yang 2000).

The results obtained with IRBP for the phylogenetic position of *Tarsius* corroborate those obtained through the study of other DNA sequences (Goodman et al. 1998), Alu repeats (Zietkiewicz et al. 1999), SINE insertions (Schmitz et al. 2001), and composite trees (Purvis 1995). To the contrary, paleontological studies cannot distinguish between three alternative hypotheses: either *Tarsius* is sister group of strepsirrhines, or it branches before the anthropoid-strepsirrhine split, or it forms a trifurcation with anthropoids and strepsirrhines (Gregory 1910; Simpson 1945; Shoshani et al. 1996; Fleagle 1999). Alternative points of view also occur with molecular studies: Murphy et al. (2001a) analyzed about 10 kb from 18 orthologous

mitochondrial and nuclear DNA segments, and showed that *Tarsius* clusters with *Lemur* in a Prosimia clade. Either the strong support for Haplorrhini here reflects a gene sampling artifact because of a peculiar behavior of the IRBP marker, or there is a taxon-sampling artifact in the data matrix of Murphy et al. (2001a) because only one strepsirrhine (*Lemur*) is included. Moreover, analyzing their data sets after removal of mitochondrial genes, as well as those nuclear genes not sequenced for *Tarsius*, severely decreases the support for the prosimian hypothesis (BP drops from 99 to 59; analyses not shown).

Table 3. Nonparametric tests of alternative hypotheses based on IRBP nucleotides (with and without third-codon positions) and amino acids for phylogenetic position of *Tarsius*, Cebidae monophyly, and association between Cheirogaleidae and Lorisiformes[1].

Phylogenetic hypotheses evaluated	Nucleotides codon positions 1 + 2 + 3				Nucleotides codon positions 1 + 2				Amino acids			
	-ln*L*	Δ	S. E.	P_{SH}	-ln*L*	Δ	S. E.	P_{SH}	-ln*L*	Δ	S. E.	P_{SH}
1. Position of Tarsius												
Tarsius+Anthropoidea (= Haplorrhini)	12 295.94	—	—	—	5 568.88	—	—	—	5 724.25	—	—	—
Tarsius + Strepsirrhini (= Prosimia)	12 301.93	5.98	4.19	0.08	5 572.92	4.04	3.58	0.13	5 726.09	1.84	3.77	0.48
Tarsius sister-group to all other Primates	12 300.43	4.49	4.85	0.18	5 571.00	2.12	4.51	0.32	5 724.83	0.58	4.39	0.59
2. Monophyly of Cebidae	12 306.18	10.24	7.16	0.08	5 576.50	7.62	5.87	0.10	5 733.06	8.81	6.90	0.11
3. Cheirogaleidae + Lorisiforms	12 350.48	54.54	15.69	<0.01	5 600.47	31.59	12.43	0.01	5 754.44	30.19	13.36	0.02

[1] Log-likelihood of each topology (ln*L*), difference (Δ) relative to highest log-likelihood, its standard error (S.E.), and confidence probability (P_{SH}) of Kishino-Hasegawa test with Shimodaira-Hasegawa correction are given.

Phylogeny of Anthropoidea

Anthropoidea is a robustly supported clade that is subdivided into Catarrhini and Platyrrhini (Fig.1). Catarrhini is also highly supported, and defined by one diagnostic insertion of six consecutive nucleotides at positions 1111–1116 of the human IRBP. Within catarrhines, the four taxa here included cluster into a cercopithecoid and a hominoid group.

The monophyly of platyrrhines is strongly supported as well (BP = 100). This confirms the occurrence of a unique colonization event of South America by anthropoids, in agreement with Goodman et al. (1998), Canavez et al. (1999), and von Dornum & Ruvolo (1999), but in contrast with the immunological results of Bauer & Schreiber (1997). Within platyrrhines, the evolutionary relationships are less well resolved (BPs within this clade do not exceed 65), but Cebidae (sensu Goodman et al. 1998), here represented by Callithrichinae (*Callithrix*) + Cebinae (*Cebus* and *Saimiri*),

are paraphyletic. Constraining the monophyly of Cebidae involves a severe drop in log-likelihood, and yields a topology that is significantly worse relative to the best one (Table 3, marginal significance of the KH-SH test: $0.08 < P < 0.11$, depending on the characters considered). Moreover, our IRBP phylogeny does not match the traditional morphological view, which groups *Pithecia, Saimiri, Cebus,* and *Ateles* into Cebidae (Rowe 1996).

The relationships within platyrrhines have already been studied with mitochondrial (Horovitz & Meyer 1995) and nuclear (Harada et al. 1995; Schneider et al. 1996; von Dormum & Ruvolo 1999) sequences, without succeeding in resolving their phylogeny. The difficulty in resolving the evolutionary affinities between the different platyrrhine families, and the incongruence of topologies obtained with different markers, could reflect a fast radiation in this group. This is confirmed by two surveys with respectively 6700 and 6723 bp of concatenated genes (Schneider 2000; Schneider et al. 2001) that cannot solve the phylogenetic relationships between pitheciines, atelines, and cebids, because they appeared almost at the same time.

Phylogeny of Strepsirhini

Strepsirrhine primates constitute a robustly supported clade (Fig.1; BP = 99). It contains two reciprocally monophyletic groups: Lemuriformes and Lorisiformes, the monophyly of the latter being the most strongly evidenced (BP = 100 vs. 83). The monophyly of lemuriforms confirms the hypothesis of a single migration event of strepsirrhines to Madagascar, in agreement with Yoder et al. (1996a, 2003). Actually, constraining the lemuriform family Cheirogaleidae to branch with lorisiforms is a significantly worse phylogenetic alternative ($P < 0.02$; Table 3). The characters shared by cheirogaleids and lorisiforms (such as the anatomy of their ascending pharyngeal artery) thus appear to be symplesiomorphous for strepsirrhines (Yoder 1994), or to have been convergently acquired by both groups. Analysis of dental and anatomical characters showed that Cheirogaleidae clusters with Lemuriformes, but not with Lorisiformes (Marivaux et al. 2001), in agreement with the present IRBP data.

Within Lemuriformes, the phylogenetic position of the aye-aye has been controversial. Because of its morphological specializations, *Daubentonia* has sometimes been considered as a basal member of strepsirrhines (Groves 1989) and even of primates (Oxnard 1981). However, in the present study, *Daubentonia*

branches with strong support as the sister group of all other lemuriforms (BP = 99), as suggested by other molecular studies (Yoder et al. 1994, 1996; Porter et al. 1995) and karyotype comparisons (Rumpler et al. 1988). As for the remaining lemuriform families, IRBP shows that Cheirogaleidae and Lemuridae are well-defined (BP = 99–100). There is a strong signal (BP = 95) to group Indridae (*Propithecus*) with Cheirogaleidae (*Cheirogaleus* and *Microcebus*). This suggests the resolution of the trifurcation between Indridae, Lemuridae, and Cheirogaleidae shown by Porter et al. (1995) and Goodman et al. (1998). It also contrasts with the cytochrome *b* analyses of Yoder et al. (1996a,b), and the study of dental and morpho-anatomical characters of Marivaux et al. (2001) in which Cheirogaleidae and Lemuridae cluster together.

Within Lorisiformes, three nodes are strongly supported: *Nycticebus* + *Loris, Otolemur* + *Galago,* and the monophyly of Galagonidae (BP = 96–100). However, Loridae appear paraphyletic due to the basal position of *Perodicticus.* For a detailed discussion of the conflict between morphological and molecular data for the branching position of *Perodicticus,* one should refer to Yoder et al. (2001). Interesting to note, *Perodicticus* is the slowest-evolving species for IRBP (see below, and Fig.1). This might explain the instability of its phylogenetic position. Indeed, sequences that evolve faster (e.g. Galagonidae, *Loris,* or *Nycticebus*) might cluster together because of long-branch attraction (Felsenstein, 1978): faster evolutionary rates increase the probability of parallel homoplastic substitutions between these sequences, leading to the exclusion of the slower-evolving potto from Loridae.

Contribution of amino acids

The ML analysis of amino acids yields a primates subtree identical to the one inferred from nucleotides (Fig.1), with three exceptions: 1) *Varecia* is the most basal taxon within Lemuridae, as found by Yoder & Irwin (1999) with combined genetic data; 2) *Perodicticus* is involved in a trifurcation with Galagonidae and the two other Loridae (*Nycticebus* and *Loris*), illustrating the difficulty of molecular data to reveal the monophyly of slow lorises (Yoder et al. 2001); and 3) *Pithecia* branches with *Callithrix,* in disagreement with other studies on nuclear markers (e.g. von Dornum & Ruvolo, 1999). The latter point emphasizes the need of additional taxon sampling for the IRBP exon 1 within platyrrhines in order to stabilize their phylogeny.

Variations of evolutionary rates between and within clades

The TC, BL, and RRTree tests recorded marked evolutionary rate differences in IRBP sequences from different clades. Below, we report results involving taxa that are consistently identified by all three tests as evolving at a significantly contrasted rate: Perissodactyla and Lemuriformes are the slowest evolving, and Afrotheria is the fastest. The BL test indicates that slower-evolving taxa are all lemuriform and perissodactyl species, whereas faster evolving taxa are *Elephantulus rufescens* and *Procavia capensis* (confidence probability $P < 0.01$). The TC test indicates that Perissodactyla evolve significantly slower than Euarchontoglires ($P < 0.05$), and Afrotheria evolve faster than Xenarthra. Within primates, Lorisiformes appears to evolve faster than Lemuriformes, as already pointed out on the noncoding ε-globin gene by Goodman et al. (1998). Significant differences ($P < 0.01$) were also detected by RRTree tests. Perissodactyla is the slowest-evolving clade, evolving significantly slower than Paenungulata. Conversely, Paenungulata evolve faster than Perissodactyla and primates. Within primates, platyrrhines and lemuriforms evolve as fastest and slowest, respectively.

Interestingly, the TC test indicates that Platyrrhini evolve faster than Catarrhini. The deceleration of the evolutionary rate of nuclear markers inside the catarrhine clade is in agreement with the hominoid slowdown hypothesis (Britten 1986; Li & Tanimura 1987; Goodman et al. 1989; Koop et al. 1989; Bailey et al. 1991). This hypothesis needs to be confirmed by further analyses, in order to evaluate whether the slowdown indeed took place along the hominoid branches (usually termed “hominoid slowdown”) or along the catarrhine branches (as suggested by the IRBP gene), and making the term “catarrhine slowdown” more appropriate.

Molecular datings using maximum likelihood local clocks

The fossil record for some primate lineages such as strepsirrhines is rather poor, but the following chronological landmarks can be given. The stem group of primates might be represented by insectivore- like placentals that lived 63–65 million years (MY) ago (Gingerich 1984; Fleagle 1999). The Platyrrhini-Catarrhini and Cercopithecoidea-Hominoidea splits are estimated at 35 MY and 25 MY, respectively (Fleagle 1999). The oldest fossils of South American monkeys are dated at 26 MY

(Rosenberger et al. 1991), and this estimate fits well with the molecular study of Goodman et al. (1998). Nevertheless, the discovery of a fossil rodent from the beginning of the Oligocene (31.5 MY) in the south of Chile (Wyss et al. 1993) suggests that platyrrhine fossils older than 26 MY might also be found. From the fossil record, the lemuriform-lorisiform divergence is estimated to be 30–40 MY old (Gingerich 1984; Beard et al. 1988; Martin 1988, 1990). Calibrated with the fossil record, different molecular markers have been used to assess primate divergence dates. Mitochondrial genes display ages far more ancient than the paleontological ones (Arnason et al. 1998, 2000; Yoder & Yang 2000). This fact has been explained by the incompleteness of the fossil record of primates, and/or by different rates of molecular evolution between lineages and markers, e.g. faster rates in anthropoid primates (Adkins & Honeycutt 1994; Andrews et al. 1998; Andrews & Easteal 2000; Liu et al. 2001). Given these discrepancies between fossil and molecular data, we decided to evaluate the contribution of IRBP exon 1 to assess divergence times between clades.

Setting local molecular clocks

The above-mentioned IRBP rate heterogeneities between and within lineages call for the use of local molecular clocks. Indeed, the global clock hypothesis is significantly rejected for our IRBP data ($P < 0.01$; Table 4). Thus different molecular evolutionary rates are allocated by ML to those branches that were previously identified as displaying significantly contrasting rates. To that end, an independent local molecular clock is sequentially assigned to branches that connect slower-evolving taxa (Perissodactyla and Lemuriformes) and faster-evolving taxa (Afrotheria).

For each definition of a given local clock, a gain of log-likelihood is observed relative to the global clock hypothesis, and this increase is more pronounced for perissodactyls than for Afrotheria and Lemuriformes (Table 4). The different values for the local clocks (see Table 4) allow the quantification of the magnitude by which taxa evolve at contrasting rates, and indicate a 2.5–3-fold contrast between the slowest evolving clades of placentals (Perissodactyla, $r_1 = 0.52$; Lemuriformes, $r_3 = 0.63$) and among the fastest (Afrotheria, $r_2 = 1.43$). Such rate variations are not special to IRBP, as they have been evidenced for other nuclear markers and for other clades

of placentals (e.g. Huchon et al. 2000, 2002).

Relative to the tree without a clock (Fig.1), the definition of the perissodactyl local clock leads to a marginal acceptance of this hypothesis by the likelihood ratio test (P = 0.05; Table 4). The highest log-likelihood tree with local clocks is obtained after combining two independent rates (Perissodactyla and Lemuriformes; Table 4). This locally clock-like phylogram with its three clocks (r_1, r_2, and the default r_0 rate) is shown in Figure 3, and represents an explanation of the IRBP data that is not significantly different from the hypothesis without a clock (P = 0.12; Table 4). Therefore, three local clocks were introduced into the model and independently calibrated by six paleontological points (Fig.3), in order to estimate the age of different splitting events among placentals.

Table 4. Likelihood ratio tests for presence of global and local molecular clocks in IRBP gene of 38 mammals (including 25 primates) and partitioned according to three codon positions[1].

	-ln*L*	2d	*P*
No clock (cf. Fig. 1)	12 297.21	—	—
Global clock [108]	12 377.19	159.96	<0.01
r_1 = 0.52 (Perissodactyla) [105]	12 361.82	129.22	0.05
r_2 = 1.43 (Afrotheria) [105]	12 371.26	148.10	<0.01
r_3 = 0.63 (Lemuriformes) [105]	12 373.08	151.74	<0.01
r_1 = 0.51 (Perissodactyla), r_3 = 0.63 (Lemuriformes) [102]	12 356.90	119.38	0.12
r_1 = 0.54 (Perissodactyla), r_2 = 1.35 (Afrotheria) [102]	12 357.29	120.16	0.11

[1] Each line provides log-likelihood (–ln*L*) that corresponds to definition of a new local clock (i.e. individual substitution rates r_1– r_3), for a taxon, which is indicated between brackets. Twice the difference (2δ) of log-likelihoods between hypothesis with and without clock is also given. *P* values of likelihood ratio test measure significance of decrease of log-likelihood relative to hypothesis without clock. Number of degrees of freedom between models with and without clock is given in brackets. For example, in the case of global clock, it is equal to 108 = 3 (number of codon partitions) × 79 (5 GTR + 1 Γ + 73 branch length parameters of model without clock) – 3 (number of codon partitions) × 43 (5 GTR + 1 Γ + 37 branch-length parameters of global clock model). All sequences for which a local clock is not defined are assumed to evolve at same default rate r_0 = 1.00.

Molecular estimates of divergence ages

Several conclusions can be drawn from the results of the molecular cross-calibrations between the six fossil references (Table 5). First, there are three pairs of calibration points that are reciprocally compatible, given the standard errors on divergence ages (cf. Table 5): 1) primates (63 MY) vs. Perissodactyla (55 MY); 2) Paenungulata (60 MY) vs. Anthropoidea (34 MY); and 3) Catarrhini (20 MY) vs. Anthropoidea (34 MY). The compatibility of the latter two points would suggest that Paenungulata (60 MY) should be compatible with Anthropoidea (34 MY). However, whereas the paenungulate point at 60 MY estimates the Catarrhini split between 15-

20 MY (i.e. 17.5 ± 2.5; Table 5), the catarrhine point at 20 MY provides older estimates for the paenungulate split (61.5 – 75.5 MY, i.e. 68.5 ± 7.0). When other splitting events are considered, the compatible calibration points Paenungulata (60 MY) vs. Catarrhini (20 MY) and Anthropoidea (34 MY) vs. Catarrhini (20 MY) give estimates too old according to the fossil record for the splitting ages outside anthropoids (e.g. a range of 64.3 – 73.4 MY and 80.4 – 91.8 MY for the first splits, respectively, within Strepsirrhini and Haplorrhini; Table 5). Notably, the Anthropoidea (34 MY) vs. Catarrhini (20 MY) points indicate 59.5 – 64.0 MY for the diversification of Malagasy lemuriforms (Table 5), in agreement with the Bayesian relaxed molecular clock IRBP estimate of Yoder et al. (2003). In contrast, the primates (63 MY) vs. Perissodactyla (55 MY) points suggest the following range of mean divergence ages: 13.8 – 14.2 MY for Lorisiformes, 26.5 – 27.2 MY for the radiation of the Lemuroidea, 39.6 – 40.7 MY for Lemuriformes, 45.4 – 46.7 MY for Strepsirrhini, and 56.7 – 58.4 MY for Haplorrhini (Table 5). These divergence times are consistent with those proposed by Goodman et al. (1998) except for Lorisiformes and within anthropoids, but more recent than those found with mitochondrial genes.

To compare with mitochondrial results, Yoder et al. (1996a) proposed 62 MY, 54 MY, and 55 MY, respectively, for the strepsirrhine, lemuriform, and lorisiform radiations, based on the calibration of ML branch lengths of a cytochrome *b* tree. A recent Bayesian age estimate based on the latter marker also suggested a 50 – 78 MY credibility interval for the diversification of Malagasy primates (Yoder et al. 2003). The difference between cytochrome *b* and IRBP age estimates is especially important for the lorisiforms (13.8 – 14.2 vs. 55 MY, respectively). Given the possible divergence between lemuriforms and lorisiforms at 30 – 40 MY, as suggested by the fossil record (Gingerich 1984; Beard et al. 1988; Martin 1988, 1990), and the postulated occurrence of lorises in the Late Eocene/Early Oligocene (Simons et al. 1986), the mitochondrial estimate of divergence time appears older than the paleontological one, whereas the nuclear estimate is far too recent. Another study using mitochondrial proteins displayed older divergence dates (Arnason et al. 1998), e.g. 60 MY for the Catarrhini/Platyrrhini split, vs. 22.6 – 23.3 MY with our IRBP datings. These differences in the estimates derived from nuclear and mitochondrial markers might be the corollary of the greater resolving power of the nuclear DNA at deep phylogenetic levels (Springer et al. 2001), and/or might reflect the use of different dating methods in different studies.

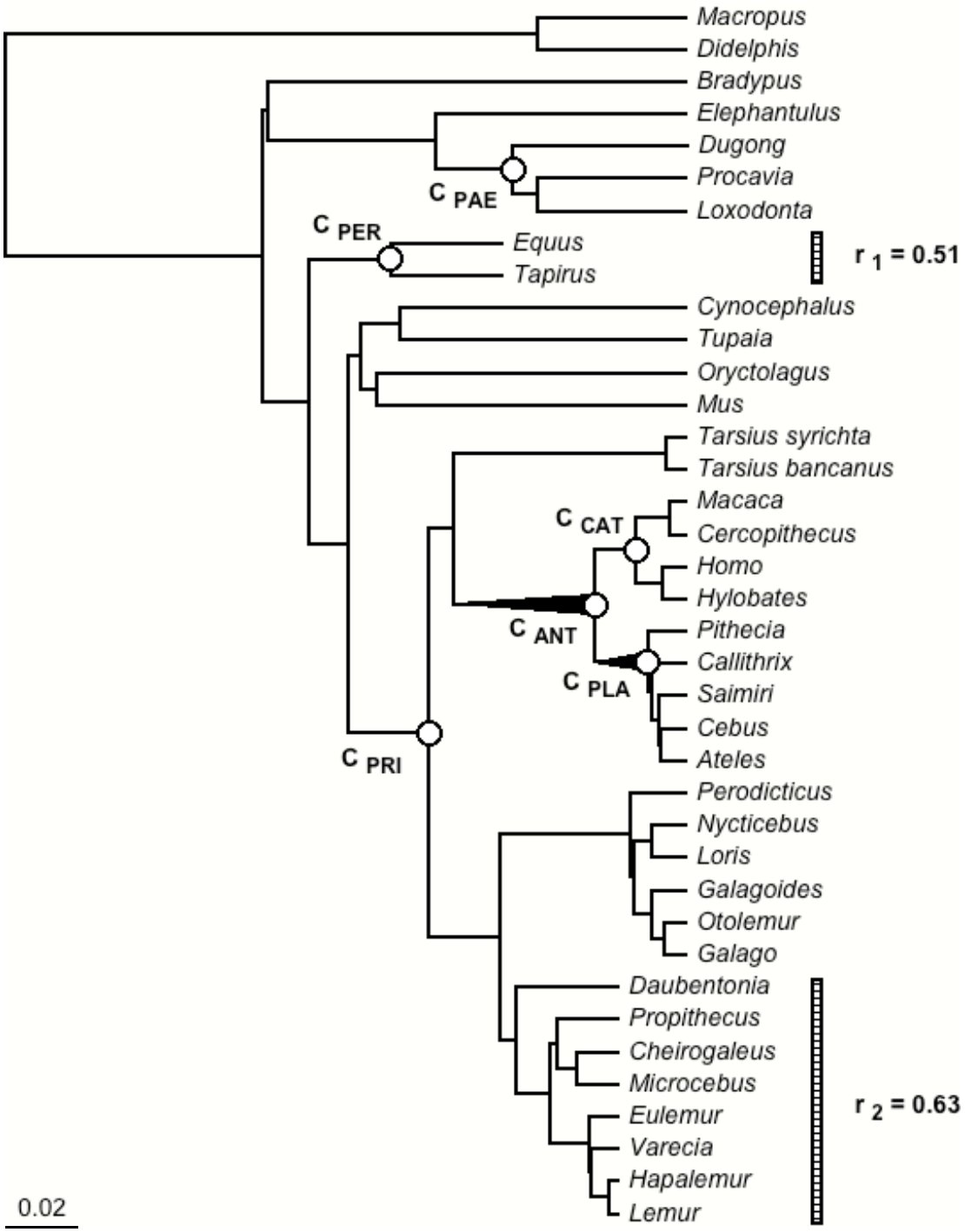

Figure 3: Local clock phylogram used to estimate divergence dates for primates. Topology is derived from highest-likelihood phylogram after exchanging positions of perissodactyls and xenarthrans _ paenungulates (see additional details in Monophyly of Primates and Their Phylogenetic Position Within Placentals). Local molecular clocks were defined for Perissodactyla ($r_1 = 0.51$) and Lemuriformes ($r_2 = 0.63$). Local clock r_1 (or r_2) was defined for all branches connecting most ancient common ancestor of perissodactyls (or lemuriformes) to two (or eight) terminal taxa, i.e. for 3 and 15 branches, respectively. Other branches evolve according to default $r_0 = 1.00$ value. Paleontological calibration points are indicated: C_{PAE} = 55–60 MY, C_{PER} = 55 MY, C_{PRI} = 63 MY, C_{ANT} = 34 MY, C_{CAT} = 20–25 MY, and C_{PLA} = 26 MY. Scale is 0.02 expected nucleotide substitution per site (i.e., 2% substitution). Thick lines correspond to branches subtending clades for which intraprimates cross-calibrations lead to incompatible results (see text for further details).

Table 5. Molecular estimates of divergence dates inferred from independent calibration points[1].

	Calibration points							
	Paenungulata		Perissodac-tyla	Primates	Anthropoi-dea	Catarrhini		Platyrrhini
	[55 Ma]	[60 Ma]	[55 Ma]	[63 Ma]	[34 Ma]	[20 Ma]	[25 Ma]	[26 Ma]
Paenungulata	—	—	43.6 (4.5)	42.3 (4.3)	**63.7 (6.5)**	68.5 (7.0)	85.6 (8.8)	118.7 (12.2)
Perissodactyla	69.4 (9.6)	75.7 (10.5)	—	**53.4 (7.4)**	80.4 (11.2)	86.4 (12.0)	108.0 (15.0)	149.8 (20.8)
Primates	81.8 (7.1)	89.3 (7.8)	**64.8 (5.6)**	—	94.8 (8.2)	101.9 (8.9)	127.4 (11.1)	176.6 (15.4)
Anthropoidea	29.4 (3.2)	**32.0 (3.5)**	23.3 (2.5)	22.6 (2.4)	—	**36.5 (4.0)**	45.7 (5.0)	63.4 (6.9)
Catarrhini	16.1 (2.3)	**17.5 (2.5)**	12.7 (1.8)	12.4 (1.7)	**18.6 (2.6)**	—	—	34.7 (4.9)
Platyrrhini	12.0 (1.6)	13.1 (1.7)	9.5 (1.2)	9.3 (1.2)	13.9 (1.8)	15.0 (2.0)	18.7 (2.4)	—
Lorisiformes	17.9 (2.3)	19.5 (2.5)	14.2 (1.8)	13.8 (1.8)	20.7 (2.7)	22.3 (2.9)	27.9 (3.6)	38.6 (5.0)
Lemuroidea	34.4 (5.4)	37.5 (5.9)	27.2 (4.3)	26.5 (4.2)	39.8 (6.3)	42.8 (6.8)	53.5 (8.4)	74.2 (11.7)
Lemuriformes	51.4 (7.0)	56.1 (7.7)	40.7 (5.6)	39.6 (5.4)	59.5 (8.2)	64.0 (8.8)	80.0 (11.0)	110.9 (15.2)
Strepsirrhini	59.0 (7.0)	64.3 (7.6)	46.7 (5.5)	45.4 (5.4)	68.3 (8.1)	73.4 (8.7)	91.8 (10.8)	127.2 (15.0)
Haplorrhini	73.7 (6.8)	80.4 (7.4)	58.4 (5.4)	56.7 (5.3)	85.4 (7.9)	91.8 (8.5)	114.7 (10.6)	159.1 (14.7)

[1] Each line corresponds to estimates of divergence ages (with standard errors [S.E.] between brackets) obtained for crown group under focus from several calibration points. Dates that are compatible, given ±1 standard error, are indicated in bold. For example, a perissodactyl calibration point at 55 Ma yields a divergence date for primates at 64.8 Ma (range ± 1 S.E., 59.2–70.4), which is compatible with 63 Ma assumed for their divergence. Estimates are also given of age of first split within strepsirrhines, lorisiforms, lemuriforms, lemuroids, and haplorrhines.

Divergence dates outside anthropoids are apparently overestimated by anthropoid and catarrhine calibration points, whereas the latter two points are underestimated by the primate and perissodactyl calibration points (Table 5). Two nonexclusive possibilities might explain this observation: 1) the IRBP substitution rate increased along the branch leading to the anthropoids; and 2) the split between modern anthropoids (C_{ANT}) is more recent than assumed, because the 34-MY-old parapithecids (i.e. the earliest fossil anthropoids) branched off deeper in the haplorrhine subtree. Within anthropoids, the platyrrhine calibration point at 26 MY consistently yields divergence ages too old for the other calibration points (e.g. the primate radiation is estimated to be 176.6 MY old). Conversely, these five other calibration points never correctly estimate the age of the platyrrhine split, leading to dates that are too recent (range, 9.3 – 18.7 MY). With regard to the far more ancient dates calculated with the platyrrhine calibration point, this might reflect large variations of the IRBP evolutionary rates after the Platyrrhini/Catarrhini split. Assuming the accuracy of the fossil estimates of 20 and 26 MY for the divergence of crown catarrhines and platyrrhines, respectively, we calculated a mean IRBP nucleotide substitution rate of 0.116%/MY for primates evolving with the default clock (Fig.3; $r_0 = 1.00$), followed by a ca. 1.5-fold increase along the anthropoid and platyrrhine ancestral branches (i.e. 0.162–0.187%/MY), and then a 4.5-fold slowdown among living platyrrhines (to 0.042%/MY). One way to attenuate the magnitude of

these substitution rate variations would be to consider that modern platyrrhines did not diversify 26 MY ago (the age of the oldest platyrrhine) but rather 14 MY ago, which would yield rates of 0.075–0.078%/MY before and after the radiation of extant platyrrhine genera, similar to those measured before and after the split of crown catarrhines (i.e. 0.078 – 0.065%/MY for a split calibrated at 20 MY). Therefore, the age of the most recent common ancestor of extant platyrrhines might be more recent than commonly assumed.

The present study stresses the fact that datings based on local molecular clocks might not be reciprocally compatible, depending on the paleontological calibration points used. This can reflect the incompleteness of the fossil record, and/or indicate that the high variability of molecular evolutionary rates in nucleotide and amino-acid sequences is not fully accommodated by local molecular clocks. Reconciliation between paleontological and molecular estimates of divergence ages in primates may benefit from the discovery of new fossils (e.g. Marivaux et al. 2001) and the sophistication of molecular dating techniques (e.g. Sanderson, 1997; Thorne et al. 1998; Huelsenbeck et al. 2000), in order to better understand how molecular evolutionary rates vary through time.

ACKNOWLEDGMENTS

We thank Franc¸ois Catzeflis for making possible our access to the Tissues Collection in Montpellier; R. Albignac, Noe¨lle Bons, M. Brack, M. Combes, Jean-Pierre Hugot, Jean-Franc¸ois Mauffrey, and Michel Tranier for collecting primate tissues; Ole Madsen and Ulfur Arnason for DNA gifts; and Wilfried W. de Jong, Stephane Ducrocq, Clark Spencer Larsen, Ole Madsen, and two anonymous reviewers for helpful comments on the manuscript. This work benefited from the financial support of the European Community TMR Network “Mammalian Phylogeny” FMRX-CT98-0221, the Génopôle Montpellier Languedoc-Roussillon, and the Action Bioinformatique Inter-EPST (2000–2002) of the CNRS. This is contribution 2003-018 of the Institut des Sciences de l’Evolution de Montpellier (UMR 5554-CNRS).

REFERENCES

Adkins RM, Honeycutt RL. (1991). Molecular phylogeny of the superorder Archonta. *Proc Natl Acad Sci USA* 88:10317–10321.

Adkins RM, Honeycutt RL. (1994). Evolution of the primate cytochrome c oxidase subunit II gene. *J Mol Evol* 38:215–231.

Allard MW, McNiff BE, Miyamoto MM. (1996). Support for interordinal eutherian relationships with emphasis on primates and their archontan relatives. *Mol Phylogenet Evol* 5:78–88.

Ammerman LK, Hillis DM. (1992). A molecular test of bats relationships: monophyly or diphily? *Syst Biol* 41:222–231.

Andrews TD, Easteal S. (2000). Evolutionary rate acceleration of cytochrome c oxidase subunit I in simian primates. *J Mol Evol* 50:562–568.

Andrews TD, Jermiin LS, Easteal S. (1998). Accelerated evolution of cytochrome b in simian primates: adaptive evolution in concert with other mitochondrial proteins? *J Mol Evol* 47:249–257.

Arnason U, Gullberg A, Janke A. (1998). Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J Mol Evol* 47:718–727.

Arnason U, Gullberg A, Burguete AS, Janke A. (2000). Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* 133:217–228.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. (2002). Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci USA* 99:8151–8156.

Bailey WJ, Fitch DHA, Tagle DA, Czelusniak J, Slightom JL, Goodman M. (1991). Molecular evolution of the _-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol Biol Evol* 8:155–184.

Bailey WJ, Slightom JL, Goodman M. (1992). Rejection of the "flying primate" hypothesis by phylogenetic evidence from the ε-globin gene. *Science* 256:86–89.

Bauer K, Schrieber A. (1997). Double invasion of Tertiary island South America by ancestral New World Monkeys? *Biol J Linn Soc* 60:1–20.

Beard KC, Dagosto M, Gebo DL, Godinot M. (1988). Interrelationships among primate higher taxa. *Nature* 331:712–714.

Borst DE, Redmond TM, Elser JE, Gonda MA, Wiggert B, Chader GJ, Nickerson JM. (1989). Interphotoreceptor retinoid-binding protein: gene characterization, proteine repeat structure, and its evolution. *J Biol Chem* 264:1115–1123.

Britten RJ. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science* 231:1393–1398.

Canavez FC, Moreira MA, Ladasky JJ, Pissinatti A, Parham P, Seuanez HN. (1999). Molecular phylogeny of new world primates (Platyrrhini) based on _2-microglobulin DNA sequences. *Mol Phylogenet Evol* 12:74–82.

Catzeflis FM. (1991). Animal tissue collections for molecular genetics and systematics. *Trends Ecol Evol* 6:168.

DeBry RW, Sagel RM. (2001). Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. *Mol Phylogenet Evol* 19:290–301.

de Jong WW. (1998). Molecules remodel the mammalian tree. *Trends Ecol Evol* 13:270–275.

Eizirik E, Murphy WJ, O'Brien SJ. (2001). Molecular dating and biogeography of the early placental mammal radiation. *J Hered* 92:212–219.

Felsenstein J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401–410.

Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.

Felsenstein J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* 22:521–565.

Fleagle JG. (1999). Primate adaptation and evolution, 2nd ed. San Diego: Academic Press.

Fong S-L, Fong W-B, Morris TA, Kedzie KM, Bridges CDB. (1990). Characterization and comparative structural features of the gene for human interstitial retinol-binding protein. *J Biol Chem* 265:3648–3653.

Garland TJ, Dickerman AW, Janis CM, Jones JA. (1993). Phylogenetic analysis of covariance by computer simulation. *Syst Biol* 42:265–292.

Gheerbrant E, Sudre J, Cappetta H. (1996). A Palaeocene proboscidean from Morocco. *Nature* 383:68–70.

Gingerich PD. (1984). Primates evolution: evidence from the fossil record, comparative morphology, and molecular biology. *Yrbk Phys Anthropol* 27:57–72.

Gingerich PD, Uhen MD. (1994). Time of origin of primates. *J Hum Evol* 27:443–445.

Goldman N, Anderson JP, Rodrigo AG. (2000). Likelihood-based tests of topologies in phylogenetics. *Syst Biol* 49:652–670.

Goodman M, Koop BF, Czelusniak J, Fitch DH, Tagle DA, Slightom JL. (1989). Molecular phylogeny of the family of apes and humans. *Genome* 31:316–335.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* 9:585–598.

Gregory WK. (1910). The orders of mammals. *Bull Am Mus Nat Hist* 27:1–524.

Groves CP. (1989). A theory of human and primate evolution. London: Oxford University Press.

Harada ML, Schneider H, Schneider MP, Sampaio I, Czelusniak J, Goodman M. (1995). DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister-grouping of Cebus and Saimiri from two unlinked nuclear genes. *Mol Phylogenet Evol* 4:331–349.

Horovitz I, Meyer A. (1995). Systematics of New World monkeys (Platyrrhini, Primates) based on 16S mitochondrial DNA sequences: a comparative analysis of different weighting methods in cladistic analysis. *Mol Phylogenet Evol* 4:448–456.

Huchon D, Catzeflis F, Douzery EJP. (2000). Variance of molecular datings, evolution of rodents, and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. *Proc R Soc Lond B* 267:393–402.

Huchon D, Madsen O, Sibbald MJJB, Ament K, Stanhope M, Catzeflis F, de Jong WW, Douzery EJP. (2002). Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol Biol Evol* 19:1053–1065.

Huelsenbeck JP, Larget B, Swofford D. (2000). A compound Poisson process for relaxing the molecular clock. *Genetics* 154: 1879–1892.

Jansa SA, Voss RS. (2000). Phylogenetic studies on didelphid marsupials I. Introduction and preliminary results from nuclear IRBP gene sequences. *J Mamm Evol* 7:43–77.

Kishino H, Hasegawa M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol* 29:170–179.

Koop BF, Tagle DA, Goodman M, Slightom JL. (1989). A molecular view of primate phylogeny and important systematic and evolutionary questions. *Mol Biol Evol* 6:580–612.

Li W-H, Tanimura M. (1987). The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93–96.

J-C, Makova KD, Adkins RM, Gibson S, Li W-H. (2001). Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. *Mol Biol Evol* 18:945–953.

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong WW, Springer MS. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610–614.

Marivaux L, Welcomme J-L, Antoine P-O, Me´tais G, Baloch IM, Benammi M, Chaimanee Y, Ducrocq S, Jaeger J-J. (2001). A fossil lemur from the Oligocene of Pakistan. *Science* 294:587–591.

Martin RD. (1988). Several steps forward for Eocene primates. *Nature* 331:660–661.

Martin RD. (1990). Primates origins and evolution: a phylogenetic reconstruction. London: Chapman & Hall.

Mercer JM, Roth VL. (2003). The effects of Cenozoic global change on squirrel phylogeny. *Science* 299:1568–1572.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. (2001a). Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady C, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. (2001b). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.

Nickerson JM, Li G-R, Lin Z-Y, Takizawa N, Si J-S, Gross EA. (1998). Structure-function relationships in the four repeats of human interphotoreceptor retinoid-binding protein (IRBP). *Mol Vis* 4:33.

Novacek MJ. (1992a). Mammalian phylogeny: shaking the tree. *Nature* 356:121–125.

Novacek MJ. (1992b). Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Syst Biol* 41:58– 73.

Novacek MJ. (1994). Morphological and molecular inrods to phylogeny. In: Thorp J, Rieppel O,

editors. Interpreting the hierarchy of nature: from systematic patterns to evolutionary process theories. San Diego: Academic Press. p 85–131.

Oxnard CE. (1981). The uniqueness of *Daubentonia*. *Am J Phys Anthropol* 54:1–21.

Philippe H. (1993). MUST: a computer package of management utilities for sequences and trees. *Nucleic Acids Res* 21:5264– 5272.

Porter CA, Sampaio I, Schneider H, Schneider MP, Czelusniak J, Goodman M. (1995). Evidence on primate phylogeny from epsilon- globin gene sequences and flanking regions. *J Mol Evol* 40:30–55.

Poux C, van Rheede T, Madsen O, de Jong WW. (2002). Sequence gaps join mice and men: phylogenetic evidence from deletions in two proteins. *Mol Biol Evol* 19:2035–2037.

Purvis A. (1995). A composite estimate of primate phylogeny. *Philos Trans R Soc Lond B* 348:405–421.

Rambaut A, Grassly NC. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci* 13:235–238.

Robinson M, Gouy M, Gautier C, Mouchiroud D. (1998). Sensitivity of the relative-rate test to taxonomic sampling. *Mol Biol Evol* 15:1091–1098.

Robinson-Rechavi M, Huchon D. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* 16:296–297.

Rosenberger AL, Hartwig WC, Wolff RG. (1991). *Szalatavus attricuspis,* an early platyrrhine primate. *Folia Primatol* 56:225–233.

Rowe N. 1996. The pictorial guide to the living primates. East Hampton, NY: Pogonias Press.

Rumpler Y, Warter S, Petter JJ, Albignac R, Dutrillaux B. (1988). Chromosomal evolution of Malagasy lemurs. 11. Phylogenetic position of *Daubentonia madagascariensis*. *Folia Primatol* 50: 124–129.

Sanderson MJ. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol* 14:1218–1231.

Schmitz J, Ohme M, Zischler H. (2001). SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. *Genetics* 157:777–784.

Schneider H. (2000). The current status of the New World monkey phylogeny. *An Acad Bras Cienc* 72:165–172.

Schneider H, Sampaio I, Harada ML, Barroso CM, Schneider MP, Czelusniak J, Goodman M. (1996). Molecular phylogeny of the New World monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences. *Am J Phys Anthropol* 100:153–179.

Schneider H, Canavez FC, Sampaio I, Moreira MAM, Tagliaro CH, Seuanez HN. (2001). Can molecular data place each neo neotropical monkey in its own branch? *Chromosoma* 109:515– 523.

Shimodaira H, Hasegawa M. (1999). Multiple comparisons of loglikelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116.

Shoshani J, Groves CP, Simons EL, Gunnell GF. (1996). Primate phylogeny: morphological vs. molecular results. *Mol Phylogenet Evol* 5:102–154.

Simons EL, Bown TM, Rasmussen DT. (1986). Discovery of two additional prosimian primate families (Omomyidae, Lorisidae) in the African Oligocene. *J Hum Evol* 15:431–437.

Simpson GG. (1945). The principles of classification and a classification of mammals. *Bull Am Mus Nat Hist* 85:1–350.

Springer M, Burk A, Kavanagh J, Waddell V, Stanhope M. (1997). The interphotoreceptor retinoid binding protein gene in therian mammals: implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc Natl Acad Sci USA* 94:13754–13759.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ. (2001.) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol Biol Evol* 18:132–143.

Stanhope MJ, Czelusniak J, Si J-S, Nickerson J, Goodman M. (1992). A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. *Mol Phylogenet Evol* 1:148–160.

Stanhope MJ, Smith MR, Waddell VG, Porter CA, Shijvi MS, Goodman M. (1996). Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J Mol Evol* 43:83–92.

Swofford DL. (2001). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b8. Sunderland, MA: Sinauer Associates.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. (1996). Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK, (editors.) Molecular systematics. Sunderland, MA: Sinauer Associates.

Szalay FS, Delson E. (1979.) Evolutionary history of the primates. New York: Academic Press.

Takezaki N, Rzhetsky A, Nei M. (1995). Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823– 833.

Thorne JL, Kishino H, Painter IS. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15:1647–1657.

von Dornum M, Ruvolo M. (1999). Phylogenetic relationships of the New World monkeys (Primates, Platyrrhini) based on nuclear G6PD DNA sequences. *Mol Phylogenet Evol* 11:459– 476.

Waddell PJ, Cao Y, Hasegawa M, Mindell DP. (1999). Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. *Syst Biol* 48:119–137.

Wagenhorst BB, Rajendran RR, Van Niel EE, Hessler RB, Bukelman A, Gonzalez-Fernandez F. (1995). Goldfish cones secrete a two-repeat interphotoreceptor retinoid-binding protein. *J Mol Evol* 41:646–656.

Whelan S, Lio P, Goldman N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet* 17:262–272.

Wyss AR, Flynn JJ, Norell MA, Swisher CC III, Charrier R, Novacek MJ, McKenna MC. (1993). South America's earliest rodent and recognition of a new interval of mammalian evolution. *Nature* 365:434–437.

Yang Z. (1996a). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 11:367–372.

Yang Z. (1996b). Maximum-likelihood models for combined analyses of multiple sequence data. *J Mol Evol* 42:587–596.

Yang Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Cabios* 13:555–556.

Yoder AD. (1994). Relative position of the Cheirogaleidae in strepsirhine phylogeny: a comparison of morphological and molecular methods and results. *Am J Phys Anthropol* 94:25–46.

Yoder AD. (1997). Back to the future: a synthesis of strepsirrhine systematics. *Evol Anthropol* 6:11–22.

Yoder AD, Irwin JA. (1999). Phylogeny of the Lemuridae: effects of character and taxon sampling on resolution of species relationships within *Eulemur*. *Cladistics* 15:351–361.

Yoder AD, Cartmill M, Ruvolo M, Smith K, Vilgalys R. (1996a). Ancient single origin for Malagasy primates. *Proc Natl Acad Sci USA* 93:5122–5126.

Yoder AD, Vilgalys R, Ruvolo M. (1996b). Molecular evolutionary dynamics of cytochrome *b* in strepsirrhine primates: the phylogenetic significance of third-position transversions. *Mol Biol Evol* 13:1339–1350.

Yoder AD, Yang Z. (2000). Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol* 17:1081–1090.

Yoder AD, Rasoloarison RM, Goodman SM, Irwin JA, Atsalis S, Ravosa MJ, Ganzhorn JU. (2000). Remarkable species diversity in malagasy mouse lemurs (Primates, *Microcebus*). *Proc Natl Acad Sci USA* 97:11325–11330.

Yoder AD, Irwin JA, Payseur BA. (2001). Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst Biol* 50:408–424.

Yoder AD, Burns MM, Zehr S, Delefosse T, Veron G, Goodman SM, Flynn JJ. (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421:734–737.

Zietkiewicz E, Richer C, Labuda D. (1999). Phylogenetic affinities of *Tarsier* in the context of primate alu repeats. *Mol Phylogenet Evol* 11:77–83.

Chapter 4

Arrival and Diversification of Caviomorph Rodents and Platyrrhine Primates in South America

Céline Poux[1,2], Pascale Chevret[1], Dorothée Huchon[3], Wilfried W. de Jong[2], and Emmanuel J. P. Douzery[1]

[1]Laboratoire de Paléontologie, Paléobiologie et Phylogénie-CC064, Institut des Sciences de l'Evolution UMR 5554/ CNRS, Université Montpellier II 34095 Montpellier, France.

[2] Department of Biochemistry 161, NCMLS, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

[3] George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel.

Systematic Biology 55:228–244. 2006

ABSTRACT

Platyrrhine primates and caviomorph rodents are clades of mammals that colonized South America during its period of isolation from the other continents, between 100 and 3 million years ago (Mya). Until now, no molecular study investigated the timing of the South American colonization by these two lineages with the same molecular data set. Using sequences from three nuclear genes (ADRA2B, vWF, and IRBP, both separate and combined) from 60 species, and eight fossil calibration constraints, we estimated the times of origin and diversification of platyrrhines and caviomorphs via a Bayesian relaxed molecular clock approach. To account for the possible effect of an accelerated rate of evolution of the IRBP gene along the branch leading to the anthropoids, we performed the datings with and without IRBP (3768 sites and 2469 sites, respectively). The time window for the colonization of South America by primates and by rodents is demarcated by the dates of origin (upper bound) and radiation (lower bound) of platyrrhines and caviomorphs. According to this approach, platyrrhine primates colonized South America between 37.0 ± 3.0 Mya (or 38.9 ± 4.0 Mya without IRBP) and 16.8 ± 2.3 (or 20.1 ± 3.3) Mya, and caviomorph rodents between 45.4 ± 4.1 (or 43.7 ± 4.8) Mya and 36.7 ± 3.7 (or 35.8 ± 4.3) Mya. Considering both the fossil record and these molecular datings, the favored scenarios are a trans-Atlantic migration of primates from Africa at the end of the Eocene or beginning of the Oligocene, and a colonization of South America by rodents during the Middle or Late Eocene. Based on our nuclear DNA data, we cannot rule out the possibility of a concomitant arrival of primates and rodents in South America. The caviomorphs radiated soon after their arrival, before the Oligocene glaciations, and these early caviomorph lineages persisted until the present. By contrast, few platyrrhine fossils are known in the Oligocene, and the present-day taxa are the result of a quite recent, Early Miocene diversification.

INTRODUCTION

South America was an isolated continent after its separation from Africa in the Cretaceous, 90 to 100 million years ago (Mya) (Smith et al. 1994), until its reconnection with North America in the Pliocene, 3 to 3.5 Mya. In the Eocene-Early

Oligocene, between 50 and 30 Mya, South America was colonized by primates and rodents, of which the extant New World platyrrhines and caviomorphs are the descendants (Flynn & Wyss 1998). No other groups of terrestrial placentals colonized South America during the same period. The times and ways of the South American colonization remain debated because of the poor fossil record available to reconstruct the evolutionary history of platyrrhines and caviomorphs fromtheir areas of origin to South America. For platyrrhines and caviomorphs alike, the most intriguing questions are (i) When and where did they diverge from their respective sister groups? (ii) When and via which route did they reach South America? (iii) Were these colonization events synchronous or not? and (iv) When did extant platyrrhines and caviomorphs begin to diversify?

Extant platyrrhines are currently divided, according to molecular studies, into three clades: Pitheciidae (e.g. *Pithecia*, the sakis), Atelidae (e.g. *Ateles*, the spider monkeys), and Cebidae (e.g. *Cebus*, the capuchins, and *Callithrix*, the marmosets) (reviewed in Schneider 2000). Resolving the relationships between these families is difficult due to the fast radiation of the platyrrhines. Thus, their branching order remains unclear in spite of a large number of studies based on mitochondrial (Horovitz & Meyer 1995) and nuclear DNA (Harada et al. 1995; Schneider et al. 1996, 2001; von Dornum & Ruvolo 1999; Schneider 2000). Old World monkeys and apes, the catarrhines from Asia and Africa, are the sister group of New World monkeys, with which they form the anthropoid clade. The origin of stem anthropoids has not yet been elucidated because the most recent discoveries revealed early anthropoid fossils in Africa (Kay et al. 1997; Gebo et al. 2000; Gunnell & Miller 2001; Seiffert et al. 2003) and in Asia (Beard et al. 1994; Jaeger et al. 1999; Marivaux et al. 2003). The Old World anthropoid fossils that share most similarities with South American primates come from the Jebel Qatrani Formation of Fayum in Egypt (Late Eocene–Early Oligocene) (Fleagle 1999). This may give some support for an African rather than Asiatic origin of the New World monkeys.

Caviomorphs belong to the rodent infraorder Hystricognathi. Like for platyrrhines, the relationships within the caviomorph clade are difficult to establish, and this is again probably due to a fast radiation (Flynn & Wyss 1998). Studies based on the mitochondrial 12S rRNA or on nuclear genes failed to give robust phylogenetic results (e.g. Nedbal et al. 1994; Catzeflis et al. 1995; Adkins et al. 2001). However, phylogenetic analyses of exon 28 of the von Willebrand factor (vWF) (Huchon &

Douzery 2001) displayed four well-defined major clades: Cavioidea (e.g. *Cavia*, the guinea pigs), Erethizontoidea (e.g. *Erethizon,* the New World porcupines), Chinchilloidea (e.g. *Chinchilla; Dinomys,* the pacaranas), and Octodontoidea (e.g. *Echimys*, the arboreal spiny rats), but their reciprocal phylogenetic affinities remained difficult to assess. According to paleontological data, hystricognaths have an Asiatic origin (Flynn et al. 1986; Bryant & McKenna, 1995; Marivaux et al. 2002), but the origin of caviomorphs could be either Asiatic (Hussain et al. 1978; Marivaux et al. 2002) or African (Lavocat 1969; Martin 1994).

Platyrrhines and caviomorphs are considered to have arrived in South America around the same paleontological time, during the Late Eocene–Early Oligocene (Hoffstetter 1972; Flynn et al. 1986). This assumption is supported by the oldest fossil findings of primates in South America at 27 Mya (Hoffstetter 1969; Rosenberger et al. 1991), and rodents at least at 31 Mya (Wyss et al. 1993). Molecular datings may help to demarcate the periods of possible colonization by estimating the time windows between the divergence of platyrrhines and caviomorphs from their respective Old World sister groups and the subsequent radiations of extant New World monkeys and rodents. Various molecular dating studies have already attempted to assess the times of origin and diversification of platyrrhines and/or caviomorphs, as summarized in Table 1, but the results are quite inconclusive.

Not only the concomitant or independent arrival of primates and rodents in the New World remains an open question, but also the possible routes and means of colonization have to be evaluated. Various biogeographical scenarios have been proposed. The most popular hypothesis is, until now, migration by rafting via floating islands from Africa to South America (Hoffstetter 1972; Houle 1999). Alternative explanations for the presence of endemic rodents and primates in South America are Gondwanan vicariance or land-mediated dispersal (Arnason et al. 2000), but both would require an unrealistically early diversification of primates and rodents, even preceding their appearance during the Late Cretaceous (e.g. Springer et al. 2003).

Defining the time of colonization of “island” areas by a given taxonomic group can be subject to different interpretations when using molecular data. It has been described either as the age of divergence of the studied clade from its mainland sister group (e.g. Vences et al. 2003; Nagy et al. 2003) or as the time of initial diversification into descending lineages on the “island” (e.g. Yoder et al. 1996, 2003;

Groombridge et al. 2002; Montgelard et al. 2003; Ross et al. 2003). These alternatives actually represent two extreme interpretations. Indeed, both approaches would give similar dating results only if the radiation of a taxon takes place immediately after the divergence from its closest mainland relative. However, the radiation of a taxon may take place long after the initial colonization event, or offshoots from early radiations may go extinct. Moreover, the extant mainland sister group of an insular clade is not necessarily its closest mainland relative, which may have gone extinct. When using extant taxa, as is the case in molecular dating studies, any such extinction events cannot be detected. Hence, to better capture the time of colonization from molecular studies, we propose a conservative approach by provid providing a time window for possible colonization demarcated by the divergence from the closest mainland sister group as an upper (i.e. oldest) bound and the ingroup diversification as a lower (i.e. most recent) bound (see Poux et al. 2005, Fig.3, for further explanation).

With this conservative approach, we estimate in the present article the dating of origin and radiation of South American primates and rodents using separate and combined sequences from three nuclear genes, coding for the alpha 2B adrenergic receptor (ADRA2B), the von Willebrand factor (vWF), and the interphotoreceptor retinoid binding protein (IRBP). The advantage of these nuclear genes is that they have already proven to be useful in solving the phylogeny of placental mammals (e.g. Madsen et al. 2001; Murphy et al. 2001; Huchon et al. 2002). This expanded data set, combined with accepted fossil calibrations (e.g. Gatesy and O'Leary, 2001; Gheerbrant et al. 2001) and a Bayesian relaxed molecular clock dating method (Thorne et al. 1998; Kishino et al. 2001; Thorne & Kishino 2002), allows for the first time to compare simultaneously the phylogeny and the evolutionary history of primates and rodents in South America using the same orthologous markers.

Table 1. Summary of molecular datings for platyrrhine and caviomorph origins and diversifications. Ages are in Mya (± standard deviation, when available). ∗Phiomorphs s.s. = Bathyergidae + Thryonomyidae + Petromuridae. IC: internal calibrations, i.e. within primates or rodents. EC: external calibrations, i.e. outside primates or rodents.

References	Platyrrhine / catarrhine split	Radiation of extant platyrrhines	Markers	Methods	Calibrations
Bailey et al. (1991)	27.2 - 34.2	12.6 - 15.9 [a]	Coding and non-coding nuclear DNA	Local clock based on MP branch lengths	IC
Takahata and Satta (1997)	57.5	-	Nuclear DNA	Global clock, ML method	—
Goodman et al. (1998)	-	25	Coding nuclear DNA	Local clock based on NJ branch lengths	IC
Kumar and Hedges (1998)	47.6 ± 8.3	-	Nuclear proteins	Global clock and multi-protein gamma distance	EC
Arnason et al. (1998)	60	-	Complete mitochondrial proteins	Local clock based on ML branch lengths	EC
Arnason et al. (2000)	70	-	Complete mitochondrial proteins	Local clock based on ML branch lengths	EC
Schneider (2000)	-	26	Nuclear DNA	Local clock based on branches calculated with the least-squares method	IC
Nei and Glazko (2002)	32.3 - 35.2	-	Nuclear proteins	Global clock and multi-protein gamma distance	IC / EC
Glazko and Nei (2003)	31.9 - 33.0	-	Nuclear proteins	Global clock and multi-protein gamma distance	IC / EC
Adkins et al. (2003)	32.4 – 49.6	-	Coding nuclear DNA	Local clock and rate smoothing	IC
Schrago and Russo (2003)	32.8 - 41.9	-	Complete mitochondrial genes and proteins	Global clock and Bayesian relaxed clock	IC
Hasegawa et al. [c] (2003)	37.5 ± 3.1	Around 17 [a, b]	Nuclear and mitochondrial DNA	Bayesian relaxed clock	EC
Yang and Yoder (2003)	53.3 - 61.1	-	Coding mitochondrial DNA	ML local clock and Bayesian relaxed clock	IC / EC
Yoder and Yang (2004)	43.5 - 55.7	-	Coding mitochondrial DNA and coding / non-coding nuclear DNA	Bayesian relaxed clock	IC / EC
	Caviomorphs / phiomorphs s.s.*	Radiation of extant caviomorphs	DNA marker	Methods	
Nebdal et al. (1994)	33 - 39	-	Mitochondrial rRNA	Global clock	IC
Huchon et al. (2000)	-	27.7 - 51	Nuclear proteins	Global clock	IC / EC
Huchon and Douzery (2001)	43 - 54	-	Coding nuclear proteins	ML local clock	IC
Mouchaty et al. (2001)	85		Complete mitochondrial proteins	Local clock based on ML branch lengths	EC
Springer et al. (2003)	31 - 46	-	Nuclear and mitochondrial DNA	Bayesian relaxed clock	IC / EC
Adkins et al. (2003)	34.7 - 57.1	-	Coding nuclear DNA	Local clock and rate smoothing	EC
Hasegawa et al. [c] (2003)	Around 40 [b]	34.4 ± 1.6 [d]	Nuclear and mitochondrial DNA	Bayesian relaxed clock	IC / EC

[a]Platyrrhine radiation age based on two species only; this date might be too young if the most basal clade is not represented.
[b]Ages deduced from the relaxed tree displayed in the article.
[c]The data set used in Hasegawa et al. (2003) is from Nikaido et al. (2001) and Murphy et al. (2001).
[d]This node was constrained to be younger than 37 Mya.

MATERIAL AND METHODS

Choice of genes

To estimate phylogeny and divergence times, the nuclear genes for ADRA2B (intronless), vWF (exon 28), and IRBP (exon 1) were chosen for the following reasons. (i) A large number of sequences was already available, especially within rodents, and these genes have been shown to contain phylogenetic information within and between mammalian orders (e.g. Madsen et al. 2001; Murphy et al. 2001; Huchon et al. 2002). (ii) The sequenced parts of these genes are of similar lengths (around 1.2 to 1.3 kb) and have comparable numbers of variable sites, which favors their equal contribution to a combined analysis. (iii) Nuclear genes have been shown to perform better than mitochondrial markers at the phylogenetic level we are interested in (Springer et al. 2001). (iv) The three genes ADRA2B, vWF, and IRBP are not genetically linked; their location is variable, on chromosomes 2, 6, and 14 in *Mus*; chromosomes 3, 4, and 16 in *Rattus*; and chromosomes 2, 12, and 10 in *Homo*. (v) The proteins they encode do not display any biological interaction or functional relation: ADRA2B is an adrenergic receptor mainly expressed in the kidneys, the vWF protein is a blood coagulation factor, and IRBP is located in the matrix of the retina. These properties allow us to combine the three genes to obtain a more reliable estimation of the phylogenies and datings, because it provides a longer data set in which the potential influence of any contrasting evolutionary properties of each individual gene is moderated.

Taxon sampling

For each gene, 60 mammalian species were included in our study, as presented in Table 2, selected on the basis of the following criteria. (i) The species should represent all placental mammalian orders, as well as two divergent marsupials as outgroups. (ii) The sampling should reflect the diversity of primate and rodent taxa; sequences fromprimates (8 for ADRA2B and 11 for vWF) and three rodents (for ADRA2B and IRBP) were therefore newly determined to include all families and/or superfamilies, with a broad representation of platyrrhines and caviomorphs. (iii)

Species should be included that enable the use of paleontological calibration constraints from various lineages, thus minimizing the dependence of the results upon a single fossil reference (Soltis et al. 2002; Douzery et al. 2003).

DNA amplification and sequencing

Newly determined primate sequences were obtained for the partial exon 28 of the vWF gene and for the ADRA2B gene. PCR reactions on the vWF gene provided two overlapping products using the primer pairs V1 (5'-TGTCAACCTCAC-CTGTGAAGCCTG-3')/ W4 (5'-TTGTTTTCAGGGGCCTGCTT-3') and V2 (5'-C-CCTCAGAGCTGCGGCGCAT-3') / W1 (5'-TGCAGGACCAGGTCAGGAGCC-TCTC-3'), and a program of 29 cycles of 20 s at 94◦C, 30 s at 47◦C, and 2 min at 68◦C, and one final cycle of 10 min at 68◦C. The PCR products were purified on Ultrafree-DA Amicon columns (Millipore) and concentrated on Microcon filterable columns (Millipore). Manual sequencing was conducted with the dideoxy chain termination method with α33P-ddNTP and the Thermo Sequenase Cycle Sequencing Kit (Amersham) on both strands, using the external primers just mentioned plus the internal primers V30 (5'-AAMTCCRTGGTTCTGGAYGTGG-3') and V40 (5'-GAGAAGCAGGCCCCNGAGAACAAGG- 3'). The almost complete coding region of the ADRA2B gene was amplified with the primers ADRA2BFOR (5'-ASCCCTACTCNGTGCAGGCNACNG-3') and ADRA2 BREV (5'-CTGTTGC-AGTAGCCDATCCARAARAARAAYTG-3'). The program was 2 min at 94◦C, 35 cycles of 1 min at 94◦C, 1 min at 54◦C, and 1 min 30 s at 72◦C, and finally 10 min at 72◦C. The PCR products were reamplified when necessary, and subsequently sequenced with the Big Dye Terminator Sequencing Kit (Applied Biosystems) and the PCR primers above, in combination with the two internal primers ADRA2B5-2 (5'-GCARGTAVACNAGRATCATG-3') and ADRA2B3-2 (5'-ATCATGATYCTN-GTBTACYTGC-3'). For the newly determined rodent ADRA2B sequences, the same protocol as given for primates was used, whereas the amplification and sequencing procedure for rodent IRBP followed Huchon et al. (2002). Species names and accession numbers of the newly determined sequences are given in Table 2.

Table 2. Taxonomic sampling and accession numbers for the three nuclear genes. Newly determined sequences are indicated (∗).

	Species	ADRA2B	vWF	IRBP
Placentalia				
RODENTIA				
Sciurognathi				
Muridae	*Mus musculus*	M94583	U27810	AF126968
	Rattus norvegicus	M32061	AJ224673	AJ429134
	Tachyoryctes sp.	AJ427264	AJ402713	AJ427231
Dipodidae	*Dipus sagitta*	AJ427263	AJ224665	AJ427232
Heteromyidae	*Dipodomys merriami*	AJ427261	AJ427226	AJ427233
Geomyidae	*Thomomys talpoides*	AJ427262	AJ427227	AJ427234
Gliridae	*Glis glis*	AJ427258	AJ224668	AJ427235
	Dryomys nitedula	AJ427257	AJ224666	AJ427236
Sciuridae	*Marmota monax*	AJ427255	AJ224671	AJ427237
Aplodontidae	*Aplodontia rufa*	AJ427256	AJ224662	AJ427238
Castoridae	*Castor canadensis*	AJ427260	AJ427228	AJ427239
Anomaluridae	*Anomalurus sp.*	AJ427259	AJ427229	AJ427230
Ctenodactylidae	*Massoutiera mzabi*	AJ427265	AJ238388	AJ427242
Hystricognathi				
Thryonomyidae	*Thryonomys swinderianus*	AJ427267	AJ224674	AJ427243
Petromuridae	*Petromus typicus*	AJ427268	AJ251144	AJ427244
Bathyergidae	*Bathyergus suillus*	AJ427252	AJ238384	AJ427251
Hystricidae	*Trichys fasciculata*	AJ427266	AJ224675	AJ427245
Chinchillidae	*Chinchilla lanigera*	AJ427271	AJ238385	AJ427246
Dinomyidae	*Dinomys branickii*	AM050859 *	AJ251145	AM050862 *
Echimyidae	*Echimys chrysurus*	AJ427269	AJ251141	AJ427247
Octodontidae	*Octodon lunatus*	AM050860 *	AJ238386	AM050863 *
Caviidae	*Cavia porcellus*	AJ271336	AJ224663	AJ427248
Agoutidae	*Agouti paca*	AM050861 *	AJ251136	AM050864 *
Erethizontidae	*Erethizon dorsatum*	AJ427270	AJ251135	AJ427249
LAGOMORPHA				
Leporidae	*Oryctolagus cuniculus*	Y15946	U31618	Z11812
	Lepus crawshayi	AJ427254	AJ224669	AJ427250
Ochotonidae	*Ochotona princeps*	AJ427253	AJ224672	AY057832
PRIMATES				
Hominidae	*Homo sapiens*	M34041	X06828	J05253
Hylobatidae	*Hylobates lar*	AM050851 *	AJ410300*	AJ313478
Cercopithecidae	*Macaca mulatta*	AM050852 *	AJ410302*	AJ313476
	Cercopithecus solatus	AM050853 *	AJ410301*	AJ313477
Cebidae	*Callithrix jacchus*	AM050856 *	AJ410299*	AJ313472
	Cebus apella	AM050854 *	AJ410297*	AJ313473
Atelidae	*Ateles sp.* [1]	AM050855 *	AF061059	AJ313474
Pitheciidae	*Pithecia pithecia*	AM050857 *	AJ410298*	AJ313475
Tarsiidae	*Tarsius bancanus*	AJ891081	AJ410296*	AF271423
Lemuridae	*Lemur catta*	AJ891067	AJ410292*	AJ313470
Indridae	*Propithecus verreauxi*	AJ891076	AJ410294*	AJ313471
Cheirogalidae	*Microcebus murinus*	AM050858 *	AJ410295*	AJ313469
Loridae	*Nycticebus coucang*	AJ251186	AJ410291*	AJ313467
DERMOPTERA				
Cynocephalidae	*Cynocephalus variegatus*	AJ251182	U31606	Z11807
SCANDENTIA				
Tupaiidae	*Tupaia sp.* [2]	AJ251187	U31624	Z11808
PHOLIDOTA				
Manidae	*Manis sp.*	AJ251185	U97535	AF025389
CARNIVORA				
Felidae	*Felis catus*	AJ251174	U31613	Z11811
PERISSODACTYLA				
Equidae	*Equus* sp.[3]	Y15945	U31610	U48710

	Species	ADRA2B	vWF	IRBP
Tapiridae/Rhinocerotidae	*Tapirus* sp. / *Ceratotherium*[4]	AJ315939	U31604	AF179294
CETARTIODACTYLA				
Bovidae	*Bos taurus*	Y15944	X63820	M20748
Physeteridae	*Physeter catodon*	AJ427417	AF108834	U50818
Hippoptamidae	*Hippopotamus amphibius*	AJ251178	AF108832	AF108837
Camelidae	*Lama* sp.[5]	AJ315941	AF108835	AF108836
Suidae	*Sus scrofa*	AJ251177	S78431	U48588
CHIROPTERA				
Pteropodidae	*Cynopterus sphinx*	AJ251181	U31605	U48709
Megadermatidae	*Megaderma lyra*	AF337537	U31616	AY057833
Phyllostomidae	*Tonatia bidens*	AF337541	U31622	Z11810
EULIPOTYPHLA				
Erinaceidae	*Erinaceus europaeus*	Y12521	U97536	AF025390
PROBOSCIDEA				
Elephantidae	*Elephas / Loxodonta*[6]	Y12525	U31611	U48711
SIRENIA				
Dugongidae	*Dugong dugon*	Y15947	U31608	U48583
HYRACOIDEA				
Procaviidae	*Procavia capensis*	Y12523	U31619	U48586
TUBULIDENTATA				
Orycteropodidae	*Orycteropus afer*	Y12522	U31617	U48712
XENARTHRA				
Bradypodidae	*Bradypus tridactylus*	AJ251179	U31603	U48708
Marsupialia				
Didelphimorphia	*Didelphis* sp.[7]	Y15943	AF226848	Z11814
Diprotodontia	*Macropus* sp.[8]	AJ251183	AJ224670	AJ429135

[a] *Ateles paniscus* (IRBP, ADRA2B) or *A. belzebuth* (vWF).
[b] *Tupaia tana* (ADRA2B) or *T. glis* (vWF, IRBP).
[c] *Equus asinus* (vWF) or *E. caballus* (IRBP, ADRA2B).
[d] *Tapirus pinchaque* (IRBP) or *T. terrestris* (ADRA2B), and *Ceratotherium simum* (vWF).
[e] *Lama glama* (vWF, IRBP) or *L. pacos* (ADRA2B).
[f] *Elephas maximus* (vWF, ADRA2B) or *Loxodonta africana* (IRBP).
[g] *Didelphis virginiana* (vWF, IRBP) or *D. marsupialis* (ADRA2B).
[h] *Macropus giganteus* (vWF, IRBP) or *M. rufus* (ADRA2B).

Phylogenetic analyses

The sequences of the three nuclear exons were highly conserved in length and easily aligned by hand with the ED editor of the MUST package, version 2000 (Philippe 1993). Nonsequenced positions and gaps were coded as missing data. Amino acid repeats and sites not sequenced or gapped in more than 25% of the taxa were excluded from analysis. The final alignment included 1188 sites for ADRA2B, 1281 sites for vWF, and 1299 sites for IRBP. Phylogenetic reconstructions were performed on the complete DNA data set by maximum likelihood (ML) withPAUP*(version 4 beta 10) (Swofford 1999) and by Bayesian analyses with MrBayes (version 3.0 beta 4) (Ronquist & Huelsenbeck 2003).

The ML assumptions were chosen after running ModelTest 3.5 (Posada & Crandall 1998). The Akaike information criterion (AIC) applied to the complete data set showed that the best fitting model of DNAsequence evolution was general time reversible (GTR) with a gamma distribution (Γ) coupled to a fraction of invariable (INV) sites to describe the substitution rate heterogeneities among sites (Yang 1996). Maximum likelihood parameters and best topology were estimated by PAUP∗ using a loop approach on the concatenated ADRA2B + vWF + IRBP genes. First, the ML parameters were optimized on an NJ topology derived from ML distances obtained using the selected model from ModelTest. Second, an ML heuristic search was conducted with tree bisection-reconnection (TBR) branch swapping to identify the optimal tree. Then, the likelihood parameters were reestimated on the new topology, and a new heuristic tree search was run under these new parameters. This loop procedure was performed until stabilization of both topology and parameters. The stability of the nodes was estimated by bootstrap (Felsenstein 1985), with 100 replicates of heuristic searches. For each replicate, we used NJ starting trees, with ML parameters identically set to their optimal value previously estimated during the loop procedure, and TBR branch swapping limited to 1000 rearrangements.

For the Bayesian inference, the nuclear DNA data set was divided into 9 partitions (3 codon positions × 3 genes). According to the best fitting models proposed by ModelTest 3.5 for each codon position separately, two different models were combined: one using GTR+Γ+INV parameters independently estimated for the two first codon position of the three exons, and one using GTR+Γ parameters independently estimated for the third codon positions of each gene. Posterior distributions were approximated by a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) technique. Five incrementally heated chains were sampled every 20 generations during 1,000,000 generations ("short-run," burn-in on the first 1500 trees) for the first analysis, and every 100 generations during 10,000,000 generations ("long-run," burn in on the first 50,000 trees) for the second. This second analysis was conducted to verify that a 10-fold increase of the number of MCMCMC generations did not affect the phylogenetic conclusions. We used Dirichlet priors for base frequencies (1,1,1,1) and for GTR parameters (1,1,1,1,1) scaled to the G-T rate, a uniform (0.05,50.00) prior for the Γ shape, and an exponential (10.0) prior for branch lengths. All topologies were a priori equally probable.

The DNA character matrix and tree are available from TreeBASE under study accession number S1389 and matrix accession number M2481.

Molecular dating analyses

The molecular dating analyses were performed according to the Bayesian relaxed molecular clock approach (Thorne et al. 1998; Kishino et al. 2001), using the MULTIDIVTIME package (Thorne & Kishino 2002). First, the best fitting parameters for each of the three codon positions in the ADRA2B, vWF, and IRBP genes (9 partitions in total) were calculated via PAML (Yang 1997), using the F84 model (because more complex models are not incorporated in the ESTBRANCHES program) and five discrete gamma categories. These parameters were then entered in the ESTBRANCHES program to calculate the branch lengths of the rooted ML tree, shown in Figure 1, and their variance-covariance matrix under each of the 9 partitions. Second, a priori knowledge was incorporated about the gamma distributions of (i) the root age, (ii) the substitution rate at the root, and (iii) the substitution rate autocorrelation along branches. These priors were specified as means and standard deviations (SD) of the three distributions according to the MULTIDIVTIME guidelines. The posterior distributions of node times were approximated through MCMC runs using the MULTIDIVTIME program. To check that the data provide significant dating information, we computed both prior and posterior divergence time distributions. Posterior distributions were also computed twice for the combined data set using all fossil calibrations, starting the MCMC runs from different initial values. The calculations have been done on each gene independently as well as on the combined data set, using the topology obtained from the concatenated alignment (Fig.1). Furthermore, the priors were the same for all 9 partitions, except for the substitution rate at the root, for which priors were recalculated for each gene separately.

For fossil calibrations we selected eight time constraints that have been used already widely in molecular dating studies (e.g. Douady & Douzery 2003; Douzery et al. 2003; Springer et al. 2003). As calibration constraints outside the primate and rodent clades we took the diversification age of Paenungulata (54 to 65 Mya; Gheerbrant et al. 2001), Perissodactyla (54 to 58 Mya; Garland et al. 1993), Cetartiodactyla (55 to 65 Mya; Gatesy & O'Leary 2001) and Lagomorpha (minimum

age, i.e. lower boundary of 37 Mya; identification of ochotonids since Late Eocene; McKenna & Bell 1997). Within the rodents we used the split *Glis/Dryomys* (minimum age of 28.5 Mya; identification of first Glirinae in Late Oligocene; Hartenberger 1994) and the split *Aplodontia/Marmota* (minimum age of 37 Mya; identification of first Sciuridae in Late Eocene; McKenna & Bell 1997). Within primates we used the basal primate radiation (63 to 90 Mya; Martin, 1993; Gingerich & Uhen 1994; Tavaré et al. 2002) and the Cercopithecoidea/ Hominoidea divergence (25 to 35 Mya; Shoshani et al. 1996; Fleagle 1999).

Additional analyses were conducted to assess the impact on dating estimates of (i) a reduced taxon sampling among platyrrhines, and (ii) the use of taxonomic chimeras. Present-day platyrrhines and caviomorphs are divided into three and four well-defined clades, respectively. Our study comprises only four species as representatives of platyrrhines (one Atelidae, one Pitheciidae, and two Cebidae), versus seven for caviomorphs (one Erethizontoidea, two Cavioidea, two Chinchilloidea, and two Octodontoidea). To assess the potential impact on dating estimates of having a smaller taxon sampling within primates, we reestimated divergence times after removing three of the caviomorphs (*Cavia, Dinomys, Octodon*), in order to reach a minimal taxon sampling (each of the four superfamilies of caviomorphs is then represented by only one species).

Moreover, in order to evaluate the impact on dating estimates of the use of taxonomic chimeras among placentals, we constructed a 67-taxon supermatrix of characters without chimera. In this matrix, all chimera of the previous analysis were disassembled and replaced by two or three sequences depending on the number of sequences used to construct the chimera. Missing data were incorporated when a given species was not represented for a given gene. In other words, we included *Ateles paniscus* and *A. belzebuth, Tupaia tana* and *T. glis, Equus asinus* and *E. caballus, Tapirus pinchaque, T. terrestris*, and *Ceratotherium simum, Lama glama* and *L. pacos, Loxodonta africana* and *Elephas maximus*, respectively, instead of *Ateles* sp., *Tupaia* sp., *Equus* sp., Ceratomorpha, *Lama* sp., and Elephantidae. When a species was not scored for a given character partition during the Bayesian relaxed molecular clock analysis, it was automatically removed by the ESTBRANCHES program for the branch lengths computation under that partition.

Statistical tests of compatibility of calibration constraints

The reciprocal compatibility of the eight calibration constraints was analyzed by repeating the dating calculations after removal of each one of the calibration constraints in turn. To assess the stability of the dating results when removing a specific calibration point, we performed an analysis of variance (ANOVA) and an H.S.D. Tukey's test (Tukey's Honestly Significantly Difference test) with the program SPSS 10.0.7 for Windows (SPSS Inc., Chicago, IL). The tests were performed on the difference between the dates calculated with all the calibration constraints and the dates found after removal of a specific calibration point.

RESULTS

Phylogenetic relationships

The ML and long-run Bayesian analysis of the three concatenated genes ADRA2B, vWF, and IRBP yield identical topologies. The four major mammalian clades (Murphy et al. 2001) are recovered: Afrotheria (ML bootstrap percentage [BP] = 100; posterior probability [PP]= 1.00), Xenarthra (only represented here by *Bradypus*), Laurasiatheria (BP = 74; PP = 1.00), and Euarchontoglires (BP = 66; PP = 1.00). Boreoeutheria, comprising Euarchontoglires and Laurasiatheria, is also strongly supported.Within primates and rodents the monophyly of platyrrhines and caviomorphs, respectively, get maximal support (BP = 100; PP = 1.00). The anthropoid clade that joins platyrrhines to catarrhines also receives maximal support. Amongst the rodent nodes of direct interest, Hystricognathi and Phiomorpha s.s. (Bathyergidae plus Thryonomyoidea) get maximal support as well; only the phylogenetic relation between Caviomorpha and Phiomorpha s.s. is somewhat less supported (BP = 69 and PP = 1.00).

Very minor differences were found between the maximum posterior probability topologies of the shortrun and long-run Bayesian analyses—i.e. 1,000,000 versus 10,000,000 MCMCMC generations. For example, *Echimys* and *Octodon* are either in basal position within caviomorphs (PP = 0.51 for the shortest run) or sister group of *Chinchilla* + *Dinomys* (PP = 0.53 for the longest run). Moreover, the

interpretation of posterior probabilities in terms of node support was identical for the two Bayesian analyses. The support for the strongest nodes (PP ≥ 0.99) remained unaffected by longer runs. Posterior probabilities of weaker nodes (PP ≤ 0.98) appeared more variable, as illustrated by the *Dugong* + *Procavia* association (short run: PP = 0.79; long run: PP = 0.90). However, all nodes in our phylogenetic tree (Fig.1) that are relevant to the understanding of the South American migration and diversification of primates and rodents are well supported, providing a reliable phylogenetic framework for the assessment of divergence times.

Molecular datings based on three combined genes and using all calibration constraints

The log-likelihood of the best tree without clock constraint was $\ln L_{NO\ CLOCK} = -71{,}720.63$, against $\ln L_{CLOCK} = -72{,}690.56$ under the global clock constraint. A likelihood-ratio test significantly rejected the hypothesis of a clock-like behavior of our data: $\delta = 2 \times (\ln L_{NO\ CLOCK} - \ln L_{CLOCK}) = 1{,}939.86$; d. f. = 60; $P < 0.001$. We therefore proceeded to a relaxation of the molecular clock hypothesis through a Bayesian approach of substitution rate autocorrelation.

First of all, the combined ADRA2B, vWF, and IRBP genes contain dating signal because the prior and posterior distributions of the divergence times are markedly different, and divergence times are converging toward the same estimates when the MCMC are run from different starting states (data not shown). The follow following informations from the dating results (Table 3 and Fig.2) are most relevant for our purposes: (i) the age of the stem groups, i.e. the time at which platyrrhines and caviomorphs diverged from their closest extant sister group—catarrhines and phiomorphs, respectively— before they arrived in South America: this will be the upper (= deeper) bound of the estimated time of arrival in South America; (ii) the age of the crown groups, i.e. the time of the earliest diversification of platyrrhines and caviomorphs (here represented by extant species only), which must have occurred after their arrival in South America: this will be the lower bound of the estimated time of arrival in South America; and (iii) the time interval between these two events, which demarcates the period during which primates and rodents may have reached South America. This interval should capture with sufficient statistical significance the actual time window during which the colonization took place.

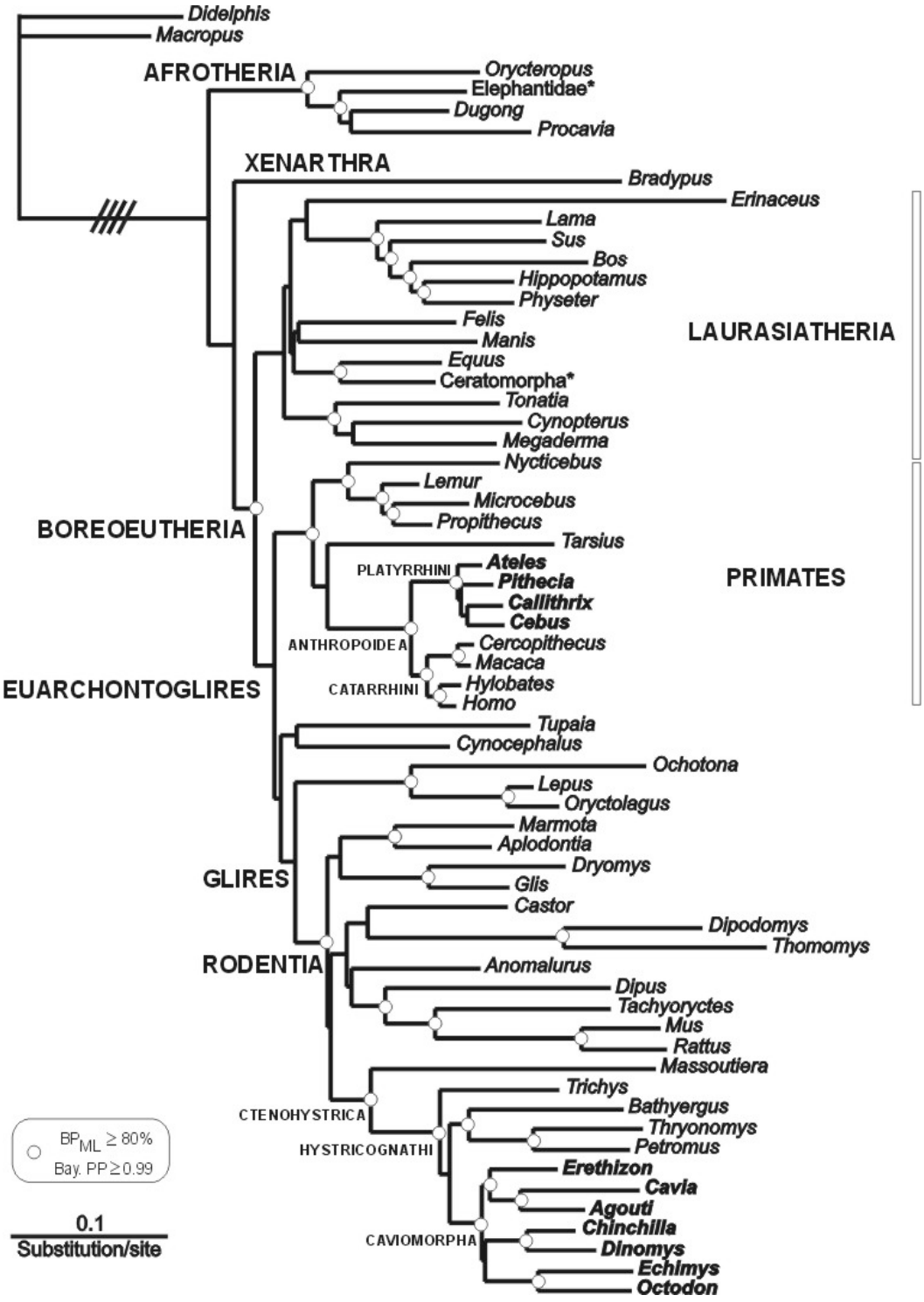

Figure 1: Phylogenetic tree reconstructed by ML analysis of the three concatenated markers ADRA2B, vWF, and IRBP (−ln*L* 71,720.63). The Bayesian analyses gave the same topology. Nodes getting a bootstrap support (BP) ≥80% and a posterior probability (PP) ≥0.99 are marked with an open circle. The length of the branch connecting the eutherians to the marsupial outgroup has been reduced four times. The names of South American primates and rodents are in bold. ∗Elephantidae and Ceratomorpha are represented in the data set by concatenated sequences from different genera (see Table 2). The names of supraordinal clades are documented in Springer et al. (2004b) and references therein. Ctenohystrica is according to Huchon et al. (2000) and Hystricognathi according to Tullberg (1899).Within primates the higher taxon names are given according to Fleagle (1999), and for Caviomorpha, see McKenna and Bell (1997).

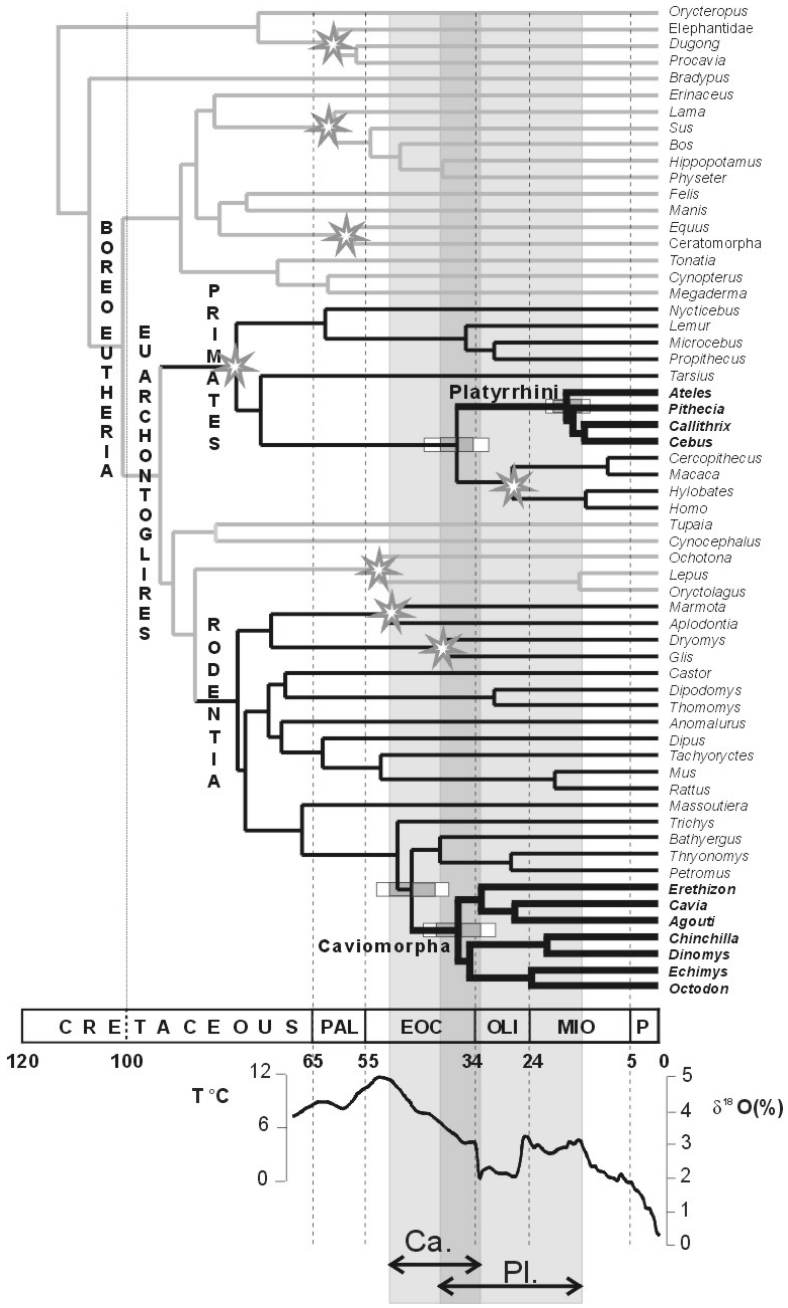

Figure 2: Chronogram showing the posterior divergence ages of placental taxa. The topology corresponds with the ML tree in Figure 1. Divergence times have been estimated from the concatenated ADRA2B, IRBP, and vWF sequences by a Bayesian relaxed molecular clock method with eight fossil calibration time constraints (nodes indicated by a star). For the nodes demarcating the period during which platyrrhines and caviomorphs may have reached South America, ±1 standard deviation and 95% credibility intervals are indicated by dark and light rectangles, respectively. The vertical gray zone spans the periods between the origin and radiation of Caviomorpha (Ca) and Platyrrhini (Pl), whereas the dark gray zone indicates the overlapping period during which Caviomorpha and Platyrrhini could have reached South America synchronously. The chronostratigraphic scale is given with absolute geological ages, and vertical dashed lines separate the epochs. PAL, EOC, OLI, MIO, and P stand for Paleocene, Eocene, Oligocene, Miocene, and PlioPleistocene, respectively. The black curve represents the variation of deep-sea δ 18O (from Zachos et al., 2001). From these values the absolute deep-sea temperature can directly be read until the Early Oligocene (34 Mya); from that period to the present the variations of deep-sea δ 18O are the result of both changes in temperature and in ice volume in Antarctica and the Northern Hemisphere. The curve shows that caviomorphs diversified before the Oligocene cooling down, whereas the platyrrhines diversified after the Oligocene, during the relatively warm first half of the Miocene.

According to our datings and their standard deviations (Table 3, all calibration constraints), the platyrrhines arrived within a time window of maximally 25.5 My, between 37.0 ± 3.0 (i.e. 34.0 to 40.0) Mya—the age of the Catarrhini/Platyrrhini divergence—and 16.8 ± 2.3 (i.e. 14.5 to 19.1) Mya—the earliest diversification of the extant platyrrhines. One should note that 25.5 My is most probably an overestimation of the time frame for the possible arrival of platyrrhines, because extant platyrrhines may not represent the first event of diversification of the platyrrhines in South America (see Discussion). By contrast, the arrival of caviomorphs in South America must have taken place in a time window of maximally 16.5My during the Middle Eocene, between 45.4 ± 4.1 Mya—the Phiomorpha/Caviomorpha split—and 36.7 ± 3.7Mya— the earliest radiation of extant caviomorphs. Taking ±1 SD into account, there is an overlap of 7.0 My (indicated by the dark gray zone in Fig.2) between the latest possible arrival time of caviomorphs (33.0 Mya) and the earliest possible arrival of platyrrhines (40.0 Mya). Based on our data, we therefore cannot rule out the possibility of a concomitant arrival of primates and rodents in South America.

Influence of individual calibration constraints

To test the possibility that individual calibration constraints may have disproportional effects on the obtained platyrrhine and caviomorph divergence time estimates, we repeated the dating analysis after removing each of the eight calibration constraints in turn (Table 3, upper part). In all instances, the posterior divergence ages are highly concordant with those obtained with the complete set of calibration constraints, with the exception of the calibration corresponding to the divergence between Cercopithecoidea and Hominoidea. When this calibration information is removed (Table 3, upper part, in bold), the date estimates for platyrrhines become different (more recent) relative to the other posterior divergence time estimations and in strong disagreement with the anthropoid fossil record (Fleagle 1999) (see Discussion). The times estimated for the caviomorphs become more recent as well, compared to the other datings, but the discrepancies are smaller than for the primates, and the results are not in disagreement with the fossil record.

Also the posterior divergence time estimate for each calibration node remains correctly recovered by all other calibrations, again with the unique exception of the Cercopithecoidea/ Hominoidea divergence (Table 3, lower part: see the diagonal).

Table 3. Posterior divergence times in Mya (±1 standard deviation) as inferred by the Bayesian relaxed molecular clock approach from the concatenated ADRA2B, IRBP, and vWF sequences. Upper part: ages pertaining to the platyrrhine and caviomorph nodes under study. Lower part: ages pertaining to the eight fossil calibration constraints.

Radiation or branching (/)	Calibration time frame [a]	All calibration constraints [b]	Removal of the following calibration constraint during the analysis [c]:							
			Paenungulata	Cetartiodactyla	Perissodactyla	Primates	Cercopith-ecoidea / Hominoidea split	Ochotona / Leporidae split	Marmota / Aplodontia split	Dryomys / Glis split
Platyrrhini radiation		16.8 ± 2.3	16.7 ± 2.2	17.0 ± 2.3	16.6 ± 2.2	16.8 ± 2.3	**14.9 ± 2.3**	16.8 ± 2.3	16.7 ± 2.3	16.8 ± 2.2
Platyrrhini / Catarrhini		37.0 ± 3.0	36.9 ± 2.9	37.4 ± 3.0	36.8 ± 2.9	37.0 ± 2.9	**32.2 ± 3.7**	37.0 ± 3.0	37.0 ± 3.0	37.0 ± 2.9
Caviomorpha radiation		36.7 ± 3.7	36.5 ± 3.6	37.2 ± 3.9	36.4 ± 3.6	36.7 ± 3.6	**35.1 ± 3.5**	36.6 ± 3.6	36.6 ± 3.7	36.6 ± 3.6
Phiomorpha s.s. / Caviomorpha		45.4 ± 4.1	45.1 ± 4.1	45.9 ± 4.1	45.0 ± 4.1	45.3 ± 4.1	**43.9 ± 4.0**	45.2 ± 4.1	45.2 ± 4.1	45.2 ± 4.1
Paenungulata	54 - 58 MYA	59.1 ± 3.0	58.4 ± 6.0							
Cetartiodactyla	55 - 65 MYA	59.4 ± 2.6		60.0 ± 3.0						
Perissodactyla	54 - 58 MYA	55.9 ± 1.1			54.6 ± 4.7					
Primates	63 - 90 MYA	77.5 ± 4.4				77.5 ± 4.5				
Cercopithecoidea / Hominoidea	25 - 35 MYA	26.8 ± 1.5					**21.4 ± 3.0**			
Ochotona / Leporidae	< 37 MYA	51.1 ± 4.8						51.0 ± 4.8		
Marmota / Aplodontia	< 37 MYA	49.9 ± 4.5							49.7 ± 4.5	
Dryomys / Glis	< 28.5 MYA	38.9 ± 4.3								38.6 ± 4.3

[a] Paleontological calibration time frames in the order in which they appear in Figure 2 (from top to bottom).
[b] Underlined are the time estimates of platyrrhine and caviomorph nodes obtained when using all calibration points.
[c] The influence of each calibration point was tested by computing divergence ages after removing that calibration point. The upper part of the table shows that only the removal of the Cercopithecoidea/Hominoidea calibration point has a significant effect, by lowering the calculated times of radiation and divergence of platyrrhines and catarrhines (in bold; ANOVA $P < 0.001$ and H.S.D. Tuckey's test $P < 0.01$). The diagonal of the table shows that seven calibration points are correctly recovered (falling within the calibration time frame as set for that point, and close to the times obtained with all calibration points) when the point itself is excluded from the constraints. However, the Cercopithecoidea/Hominoidea split (in bold) is dated more recent (21.4 Mya) than allowed by its calibration time frame (25 to 35 Mya), and also more recent than the recovered time when all calibrations are used (26.8 Mya).

This divergence now drops to 21.4 ± 3.0 Mya, whereas the time range used as calibration constraint is 25 to 35 Mya.

Molecular datings based on separate genes

To explore the contributions of the three genes in our calculations, we estimated the divergence dates from each gene separately, with and without the Cercopithecoidea/Hominoidea calibration point (Table 4). Using all calibration constraints, ADRA2B indicates that rodents and primates may have reached South America concomitantly between 44.6 to 39.2 Mya. Similarly, the vWF data set shows an approximately concomitant migration between 38.3 to 33.1 Mya. However, absolute ages estimated by ADRA2B for rodents and the New World/OldWorld primates split are around 6My deeper as compared with the vWF. Finally, IRBP gene suggests that platyrrhines originated after the radiation of caviomorphs, which would consequently exclude a concomitant migration of primates and rodents to South America.

Removing the Cercopithecoidea/Hominoidea calibration point had no effect on vWF estimates, resulted in slightly younger estimates for ADRA2B, but had the greatest impact for IRBP estimates (Table 4). In the latter case, the datings within the anthropoids become exceedingly young as compared to the fossil record. It estimates, for example, the split between Cercopithecoidea and Hominoidea at 16.5 ± 3.6 Mya, whereas the oldest Proconsulidae (hominoids) is already present in the fossil record around 25 Mya (Fleagle 1999). Relative rate tests and nonsynonymous-to-synonymous ratios determined from the IRBP sequences (results not shown) indicate that the younger dates found within anthropoids could be explained by an increase of the molecular evolutionary rate of IRBP within this clade. This assumption is also supported by Poux & Douzery (2004), who described a higher IRBP nucleotide substitution rate along the anthropoid and platyrrhine ancestral branches. Using the Cercopithecoidea/Hominoidea calibration point allowed us to reduce the impact of the IRBP high molecular evolutionary rate on the datings.

Table 4. Posterior divergence ages in Mya (± standard deviation) for the primate and rodent nodes of interest, as inferred by the Bayesian relaxed molecular clock approach from the ADRA2B, vWF, and IRBP sequences separately, and for the combined ADRA2B + vWF sequences, with or without the Cercopithecoidea/Hominoidea calibration point. In bold: impact of the removal of the Cercopithecoidea/Hominoidea calibration on IRBP dating estimates. The results with all calibration constraints but without IRBP are underlined; they are compared, in the text, with the underlined results from Table 3 corresponding to the calculations computed with all calibration points and all genes.

	ADRA2B		**vWF**		**IRBP**		**ADRA2B + vWF**	
	All calibrations	Without Cercopithecoidea / Hominoidea split	All calibrations	Without Cercopithecoidea/ Hominoidea split	All calibrations	Without Cercopithecoidea / Hominoidea split	All calibrations	Without Cercopithecoidea / Hominoidea split
Platyrrhini radiation	20.4 ± 5.0	19.1 ± 5.0	22.8 ± 4.7	22.8 ± 5.1	15. 7 ± 3.2	**12.3 ± 2.8**	20.1 ± 3.3	19.4 ± 3.5
Platyrrhini / Catarrhini	44.6 ± 6.3	40.8 ± 7.4	38.3 ± 4.5	38.3 ± 6.3	37.2 ± 4.2	**27.5 ± 4.8**	38.9 ± 4.0	37.2 ± 5.0
Caviomorpha radiation	39.2 ± 6.6	38.5 ± 6.5	33.1 ± 5.0	33.1 ± 5.0	42.8 ± 6.1	41.5 ± 5.9	35.8 ± 4.3	35.4 ± 4.4
Phiomorpha s.s. / Caviomorpha	46.6 ± 7.0	45.9 ± 6.9	40.6 ± 5.6	40.6 ± 5.5	52.4 ± 6.7	50.7 ± 6.4	43.7 ± 4.8	43.2 ± 4.9
Cercopithecoidea / Hominoidea	28.7 ± 2.6	23.4 ± 6.0	29.9 ± 2.8	30.0 ± 5.9	26.9 ± 1.7	**16.5 ±3.6**	28.7 ± 2.8	26.7 ± 4.4
Primates	77.5 ± 6.1	76.0 ± 6.3	79.4 ± 5.3	79.3 ± 5.6	76.4 ±6.2	70.7 ± 5.7	78.5 ± 4.9	77.5 ± 5.2

Molecular datings with combined ADRA2B and vWF

Considering that the faster evolution of IRBP in anthropoids might have affected the combined results as given in Table 3, we also estimated divergence times using the combined ADRA2B and vWF genes alone (Table 4). With all calibration constraints included, the results for caviomorphs become slightly younger (order of magnitude ca. 1 My) than with the three combined genes. The impact of the IRBP removal is stronger on the platyrrhine datings, and actually makes them deeper, nowplacing the platyrrhine/catarrhine split at 38.9 ± 4.0 Mya and the platyrrhine radiation at 20.1 ± 3.3 Mya. These results are 1.9 My and 3.3 My older, respectively, than found with the three genes combined. The overlap between the time frames for arrival of rodents and primates now becomes larger: 11.4 My (between 42.9 My and 31.5 My, taking ±1 SD into account) versus 7.0 My.

However, the maximum time intervals during which rodents and primates may have reached South America remain concordant with the first calculations, namely 17.0 and 26.1 My, versus 16.5 and 25.5 My, respectively. The longer lag time between origin and diversification of the extant platyrrhines compared to caviomorphs is an aspect in which the three genes consistently agree. The dating estimates of the caviomorph and platyrrhine radiations never overlap, and it does not make any difference whether or not we omit IRBP, incorporate the Cercopithecoidea/Hominoidea split in the calibration, or use genes separately or in combination (Tables 3 and 4).

We also tested with this ADRA2B + vWF data set whether we could recover the individual calibration constraints after removing each of them in turn. A reciprocal compatibility of all the calibrations points was observed, including that for the Cercopithecoidea/ Hominoidea divergence (Table 4: 26.7 ± 4.4, to be compared to Table 3). Even though the datings after IRBP removal seem to give more appropriate results within anthropoids, the differences between the dates, with or without IRBP, will not alter our overall conclusions.

Influence of taxon sampling and use of chimeras

Divergence times estimated with only one species sampled from each of the four well-defined caviomorph superfamilies—*Erethizon* (Erethizontoidea), *Agouti*

(Cavioidea), *Chinchilla* (Chinchilloidea), and *Echimys* (Octodontoidea)—lead to very similar results as compared to the datings based on the complete taxonomic set of seven species. Sampling four instead of seven caviomorphs yields the following divergence time estimates: 44.2 ± 4.0 Mya instead of 45.4 ± 4.1 Mya for the caviomorph/phiomorph split, and 36.1 ± 3.6 Mya instead of 36.7 ± 3.7 Mya for the caviomorph radiation. Extrapolating this observation to the platyrrhines, we thus can reasonably assume that the inclusion of only one species of Pitheciidae and Atelidae has only had a minor influence on the results. Moreover, given that all three platyrrhine clades were sampled, and apparently radiated in a very short time span, a reduced sampling within each family is not expected to strongly influence the results.

Apart from the problem of a reduced taxon sampling among platyrrhines, the construction of taxonomic chimeras in order to improve the nuclear gene coverage may have biased our dating estimates. Divergence times were therefore estimated from a 67-taxon supermatrix of characters; i.e. without chimeras of placental species, and using a reference topology in which the respective monophyly of *Ateles paniscus* + *A. belzebuth, Tupaia tana* + *T. glis, Equus asinus* + *E. caballus, Tapirus pinchaque* + *T. terrestris* + *Ceratotherium simum, Lama glama* + *L. pacos*, and *Loxodonta africana* + *Elephas maximus* was assumed. We observed that the dating results are essentially identical to those estimated from the original 60-taxon matrix, containing eight chimeras: 45.1 ± 4.1 Mya and 36.5 ± 3.6 Mya for the origin and radiation of caviomorphs, and 37.0 ± 2.9 Mya and 16.8 ± 2.3 Mya for the origin and radiation of platyrrhines. The finding that the use of composite taxa did not influence our dating results is in agreement with previous analyses showing the positive contribution of composite taxa in phylogeny reconstruction, as long as the species used to build chimeras are known to be monophyletic relative to the other species in the dataset (e.g. Springer et al. 2004a).

DISCUSSION

Contribution of each gene and calibration constraint

Two major difficulties of molecular dating are (i) the variation of evolutionary rate, which can be both gene-specific and lineage-specific, and (ii) the paleontological

uncertainties associated with the calibration constraints (for reviews see Bromham & Penny 2003; Graur & Martin 2004). With regard to rate variation, a global molecular clock certainly does not fit our data (see Results), in agreement with the fact that extensive rate variations have been shown in primates and rodents, both with nuclear and mitochondrial data (e.g. Liu et al. 2001; Adkins et al. 2003; Douzery et al. 2003; Poux & Douzery 2004). We therefore used the Bayesian method of Thorne et al. (1998) and Kishino et al. (2001), which is based on a probabilistic model of evolutionary rate autocorrelation, and has already been used in various animal groups (e.g. Bossuyt & Milinkovitch 2001; Hasegawa et al. 2003; Hassanin & Douzery 2003; Springer et al. 2003; Yoder et al. 2003). In this model, rates are allowed to vary over time and lineages, and rate changes along descending branches are autocorrelated according to a lognormal model. Moreover, a distinct model of nucleotide substitution can be defined for each selected gene partition (Thorne & Kishino 2002).

The three nuclear markers used in this study do not lead to concordant datings and colonization scenarios, arguing for the combination of multiple genes to obtain an averaged representation of the underlying evolutionary processes, and thus divergence times. Our results show that, even with the use of more realistic evolutionary models, strong rate variation cannot be completely taken into account (see Table 2, Cercopithecoidea/Hominoidea split). To compensate to some extent for this effect, we will use in the following biogeographic sections both the datings obtained with the three-gene and with the two-gene (ADRA2B + vWF) combinations. An unexplained observation is that the datings obtained with the three- or two-gene combinations are always more recent than the average dates based on the separate genes (Tables 3 and 4). This intriguing feature can in fact also be observed in other dating studies where the results from combined genes might be markedly different from the average dates calculated with the separated genes (e.g. Yoder & Yang 2004).

The problem of the paleontological uncertainties is reduced by Kishino et al.'s (2001) method because it handles calibration constraints as time ranges instead of time points. It has indeed been shown that the precision of the divergence time estimates is substantially enhanced when constraints are included (Kishino et al. 2001). In our analysis it only is the Cercopithecoidea/Hominoidea calibration point that has a major influence, but only on some of the anthropoid datings. Because this calibration point is located inside the anthropoid crown group, it is the only one that can counterbalance the influence of the rate acceleration in the IRBP sequences of

anthropoids. This illustrates the importance of taking calibration constraints close to the nodes to be dated, in order to reduce the influence of local deviations in the evolutionary properties of the genes involved. The Cercopithecoidea/ Hominoidea calibration point has already been used in recent molecular studies, but with quite different assigned times: e.g. Yang & Yoder (2003) used a range of 32 to 38 Mya (partially based on molecular datings), whereas Schrago & Russo (2003) and Adkins et al. (2003) used 25 Mya based on paleontological data. As this calibration point is crucial for our datings, we used a conservative range of 25 to 35 Mya, especially in the light of the fossil record and a recent molecular study (Steiper et al. 2004). It is interesting to note that in our calculations the age of the Cercopithecoidea /Hominoidea split is never older than 30.0 Mya (Tables 3 and 4). Using a range of 32 to 38 Mya could therefore lead to biased conclusions because this time frame seems too deep, or at least too narrow. The difference in the assigned age of the Cercopithecoidea/Hominoidea calibration point could explain why Schrago & Russo (2003) and Yang & Yoder (2003) obtained such strikingly different results for the platyrrhine/catarrhine split (see Table 1), even though the same methods and mitochondrial DNA data were used.

Finally, the importance of a broad taxon sampling in combination with various calibration constraints must be emphasized. This allows the breaking of long branches, and consequently a better coping with the rate variations along such branches. Unfortunately, in our case it was not possible to break the long branch leading to the anthropoids because there are no additional extant species available.

Could the colonizations of south america by primates and rodents have been synchronous or not?

The phylogenetic relationships among primates as shown in Figure 1 agree with the current consensus from nuclear DNA phylogenies (i.e. Goodman et al. 1998; Poux & Douzery 2004), and rodent relationships are congruent with those in Adkins et al. (2001), Huchon & Douzery (2001), Huchon et al. (2002), and DeBry (2003). Given that platyrrhines and caviomorphs are restricted to South America, the monophyly of both groups suggests a single colonization event for each of them. A double invasion event has been claimed for primates on the basis of antigenic determinants (Bauer & Schreiber 1997) and for rodents on the basis of paleontological evidence (e.g. Bryant

& Mc Kenna 1995), but this is not supported by our or any other molecular data. If other clades of primates and rodents reached South America during its period of isolation, they left no extant representatives.

A general agreement was found among divergence time estimates from different studies on primates and rodents (Table 1). Our estimated dates for the catarrhine/platyrrhine split (37.0 ± 3.0 Mya for the ADRA2B + vWF + IRBP analysis/38.9 ± 4.0 Mya for the ADRA2B + vWF analysis) are in perfect agreement with the results of Hasegawa et al. (2003) (37.5 ± 3.1 Mya), and around the average dates of the other studies based on either nuclear or mitochondrial genes, or a combination thereof. Similarly, our datings of the caviomorph/phiomorph split (45.4 ± 4.1 Mya/43.7 ± 4.8 Mya) fit with those of Huchon & Douzery (2001), and are slightly older than the dates obtained by Hasegawa et al. (2003) and Springer et al. (2003). However, there are three exceptions with deviating time estimates based on mitochondrial sequences. Arnason et al. (1998, 2000) estimated the catarrhine/platyrrhine (i.e. anthropoid) split at 60 to 70 Mya, whereas Yang & Yoder (2003: table 7) estimated it between 53.3 Mya and 61.1 Mya. Mouchaty et al. (2001) estimated the caviomorph/phiomorph split at 85 Mya. These discrepancies are likely to be the result of differences in the (i) methodology used for the calculations (global, local, or relaxed clocks); (ii) choice of calibration references (far from the clades under focus); and (iii) degree of accuracy and precision of these calibrations. It has indeed been suggested that the results obtained with nuclear and mitochondrial markers become consistent if appropriate methods and calibrations are used (e.g. Hasegawa et al. 2003).

Our dating estimates do not clearly advocate either a synchronous or an asynchronous colonization of South America by primates and rodents. As already mentioned, the periods of time during which primates and rodents could have reached South America overlap for 7.0 to 11.4 My, depending on the nuclear genes used. If we take into account that during this period, from 40.0/42.9 Mya (Middle Eocene) to 33.0/31.5 Mya (Early Oligocene), the geographic and environmental conditions allowing colonization may only temporarily have existed, and a more or less synchronous arrival can be conceived. In that case, the caviomorphs diverged from their sister group (45.4/43.7 Mya) some time before they reached South America and radiated (36.7/35.8 Mya) soon after their arrival. In contrast, the primates should have colonized South America just after the divergence from their sister clade (37.0/38.9

Mya), whereas the radiation of extant platyrrhines (16.8/20.1 Mya) began much later, any lineages resulting from earlier diversification now being extinct. In conclusion, if suitable conditions and opportunities for primates and rodents to reach South America have been extremely rare, perhaps only occurring from the Middle Eocene until the Early Oligocene (the dark gray zone in Fig.2), our datings may be in favor of a synchronous colonization; if suitable conditions existed repeatedly throughout the Eocene and Oligocene, our data would rather favor asynchronous colonizations.

Possible migration histories

Given that representatives of both primate and rodent orders did not reach South America before the Eocene, a land-bridge dispersal during the Late Cretaceous–Early Paleocene via the Rio Grande Rise and the Walvis Ridge, as proposed for primates by Arnason et al. (2000), can be dismissed. Various other biogeographical hypotheses remain open to explain the colonization of South America by primates and rodents. These hypotheses are based, first, on the locations of the oldest anthropoid, platyrrhine, and hystricognath fossils in the Old World (parsimoniously assuming that these locations might be the centers of origin of these clades) and, secondly, on the climatic and geographic conditions during the migration period to South America.

One hypothesis assumes an African origin for caviomorphs and platyrrhines, from phiomorph (Lavocat 1969; Martin 1994) and anthropoid stocks (Fleagle 1999), respectively, followed by a transatlantic migration. Despite the distance between the two continents during the Middle Eocene–Early Oligocene, colonization could have occurred, aided by marine currents, palaeowinds, or "stepping stone" islands along with rafts (Wyss et al. 1993; Flynn & Wyss, 1998; Houle 1999). Transoceanic dispersals have also been suggested for a variety of other taxonomic groups (de Queiroz 2005), including Africa/South America exchanges for squamates and angiosperms, and Africa/Madagascar exchanges for squamates and amphibians (e.g. Vences et al. 2003) and mammals (e.g. Yoder et al. 1996, 2003; Poux et al. 2005). The transatlantic route is the preferred hypothesis concerning platyrrhine migration for two reasons. First, fossils considered as early platyrrhines (parapithecids or proteopithecids from the Late Eocene Fayum formation in Egypt; Simons 1997; Fleagle 1999; Ross 2000) and early catarrhines (*Aegyptopithecus* from the Early Oligocene Fayum formation in Egypt; Fleagle 1999) have so far only been found in

Africa. Secondly, migration through Antarctica is unlikely for this group because at the time of platyrrhines/catarrhines divergence—at most, 37 Mya—Australia, Antarctica, and South America were no longer strongly connected, while Antarctica was, moreover, covered by ice sheets (Zachos et al. 2001).

With respect to rodents, an Asiatic origin of hystricognaths is broadly supported (Flynn et al. 1986; Bryant & McKenna 1995; Marivaux et al. 2002), implying that South American caviomorphs and their sister group, the African phiomorphs, share an Asian hystricognath ancestor (Marivaux et al. 2002). However, it is not clear whether caviomorphs and phiomorphs diverged already in Asia or after migration of their hystricognath ancestor into Africa, which leaves different dispersion routes to South America open to speculation. Indeed, the dispersal of the caviomorph ancestor to South America might have occurred from Africa (Lavocat 1969), but also from Asia, with a subsequent migration via Antarctica (Huchon & Douzery 2001) or North America (e.g. Hussain et al. 1978). A weak point of the latter two proposals is that no protocaviomorph remains have been reported from Antarctica or Australia, and the ones discovered in North America appear to have been misinterpreted (e.g. Martin 1994). However, in contrast to North America, the fossil record of Antarctica is relatively unexplored, and it has recently indeed been shown that transantarctic dispersal has been quite frequent in the southern hemisphere (Sanmartín & Ronquist 2004). Moreover, colonization via Antarctica was physically possible for rodents because South America was connected to Antarctica until 37 to 30 Mya (e.g. Barker et al. 1991; Lawver et al. 1992; Lawver & Gahagan 2003), and even though Australia became separated from Antarctica around 90 Mya, the two continents remained fairly close together until the opening of the Tasman Sea, 35 Mya (Lawver et al. 1992). During this period, the climate was still temperate, and angiosperms flourished on Antarctica. However, the sea barrier between Asia and Australia was at that time wider than the Atlantic Ocean and, because of this major problem, Houle (1999) refuted this possibility. From our study it is not possible to decide between a rodent colonization of South America via Africa or via Antarctica; the discovery of protocaviomorph fossils might shed more light on this issue.

When Did Platyrrhines and Caviomorphs Diversify in South America?

The oldest caviomorph fossil found in South America is estimated at 31 Mya and is considered to belong to the extant family Dasyproctidae (Wyss et al. 1993) or Dinomyidae (Vucetich et al. 1999). This means that at 31 Mya the caviomorphs had probably already started to diversify, and implies that the arrival of caviomorphs in South America predated the Early Oligocene. Such a view is supported by our Late Eocene dating of the caviomorph radiation (36.7±3.7/35.8±4.3 Mya) and is in agreement with the paleontological analysis of Vucetich et al. (1999). This result also implies that extant caviomorph lineages derive from early diversification events.

The oldest primate fossils from South America, *Branisella* and *Szalatavus* (27 Mya, Bolivia) (Rosenberger et al. 1991), considered as a single genus by Takai & Anaya (1996), are plesiomorph platyrrhines and have no direct relation with living platyrrhines (Fleagle 1999). The Patagonian primate fossils from the Early and Middle Miocene (21 to 14 Mya) are considered either as sister group of extant platyrrhines or as nested within this clade (Fleagle 1999). Thus, there are still paleontological uncertainties about the time of radiation of living South American primates. However, the primate fossils from the later Middle Miocene of La Venta (12 to 13 Mya, Colombia) are highly similar to modern platyrrhines (Fleagle 1999). Our dating of the platyrrhine diversification, during the Early Miocene (16.8±2.3/20.1±3.3 Mya), is concordant with the view that the La Venta fossils belong to the modern lineages, and that only the oldest of the Patagonian fossils might not belong to any of the extant families, but originated earlier. The platyrrhine diversity before the Early Miocene is quite poor either because of gaps in the fossil record or because platyrrhines did not undergo an explosive radiation as the caviomorphs did. Extinction events likely occurred in both platyrrhine and caviomorph lineages but, because of the poor diversity of platyrrhines, only one lineage resulting from the early platyrrhine radiation survived. Present-day platyrrhines would actually derive from a late diversification event during the Early Miocene.

Platyrrhine and Caviomorph Radiations and Global Climatic Changes

The dating of the platyrrhine and caviomorph arrivals and their subsequent diversification can be correlated with global climatic changes (see Fig.2) (Zachos et al. 2001). Such an influence has already been proposed for another South American mammalian clade, the xenarthrans (Delsuc et al. 2004). The diversification of extant caviomorphs started in the Late Eocene, before the beginning of the Oligocene glaciation periods. It appears that several caviomorph lineages were able to adapt to the Oligocene climate changes, and an example of such adaptation might be the evolution of hypsodont teeth among caviomorphs (Vucetich et al. 1999). In fact, around 31 to 35 Mya, hypsodont herbivores were dominating the mammalian fauna in South America (Tinguiririan fauna); these herbivores are usually considered as grazers, implying that open habitats (woodlands to savanna) and grasslands were present (Flynn & Wyss 1998). Rodents probably radiated by exploiting these and other new niches opened by climatic changes.

According to our dating, all present-day platyrrhines result from an Early Miocene diversification. This suggests that Oligocene platyrrhines underwent extinction events. Primates extinctions have frequently been explained by climatic changes, such as the general cooling down in the beginning of the Oligocene, leading to a strong decline of primate diversity in the northern continents (Gingerich 1986; Fleagle 1999). The African ancestors of neotropical monkeys were arboreal quadrupeds (Fleagle 1999), and all extant platyrrhines are still arboreal; only *Branisella*, the first recorded primate in South America (27 Mya), has been suggested to be semiterrestrial and its dentition adapted to abrasive food (Takai et al. 2000). It thus seems likely that Oligocene glaciations (26.5 to 34.0 Mya), resulting in the transformation of forests into open areas, may explain why early platyrrhine lineages are semi-terrestrial lineages that are not representative of the extant diversity. McKenna & Bell (1997) recognized only one genus of primates in the Oligocene and five in the Early Miocene, as compared to 16 caviomorph genera in the Oligocene and 25 in the Early Miocene. These differences in number of genera may illustrate the fact that the climatic conditions were probably not appropriate to allow primates to realize a similar explosive radiation as caviomorphs did.

CONCLUSIONS

This study is the first one comparing the colonization histories of platyrrhine primates and caviomorph rodents using the same set of (nuclear) genes. Considering both the fossil record and our molecular dating estimates, the most plausible scenario for primates suggests a transatlantic migration at the end of the Eocene, followed by the extinction of all but one of the few earlier diverging lineages, and the radiation of extant platyrrhines during the Early Miocene. Our results also show that the arrival of rodents and primates in South America might have been contemporaneous. However, in contrast to platyrrhines, representatives of the early diversification of caviomorphs, which occurred before the Oligocene glaciations, survive until the present. The absence of fossil information about caviomorphs outside South America, which could have been combined with our molecular data, allows only speculations about their migration history. A better understanding of this open biogeographical question awaits the discovery of new fossils, phylogenetically close to the caviomorphs.

ACKNOWLEDGMENTS

We thank Ole Madsen for help and advice with the ADRA2B sequencing and for his great support. This work would not have been possible without the essential contribution of François Catzeflis (curator of the tissue collection of the Institut des Sciences de l'Evolution de Montpellier), and of all tissue collectors: M. Brack, J.-P. Hugot, J.-F. Mauffrey, Faune Sauvage (EDF-CNEH), M. J. Stanhope, Mr. Combes, R. Albignac, M. Tranier, O. Madsen, and N. Bons. D.H. thanks Christopher Bonar, Tammie Bettinger, and the Cleveland Metroparks Zoo for providing the samples of *Dinomys*. We wish to thank two anonymous reviewers, R. DeBry, and R.D.M. Page, for their helpful comments to improve the present paper. This work was supported by the TMRNetwork "Mammalian Phylogeny" (contract FMRX-CT98-022) of the European Community, and the "ACI Informatique-Mathématique-Physique en Biologie Moléculaire (ACI IMP-Bio)." This publication is contribution number EPML-009 of the Equipe-Projet multi-laboratoires CNRS-STIC "Méthodes informatiques pour la biologie moléculaire," and number 2005-081 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554–CNRS).

REFERENCES

Adkins, R. M., E. L. Gelke, D. Rowe, and R. L. Honeycutt. (2001). Molecular phylogeny and divergence time estimates for major rodent groups: Evidence from multiple genes. *Mol. Biol. Evol.* 18:777–791.

Adkins, R. M., A. H. Walton, and R. L. Honeycutt. (2003). Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol. Phylogenet. Evol.* 26:409–420.

Arnason, U., A. Gullberg, A. S. Burguete, and A. Janke. (2000). Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* 133:217–228.

Arnason, U., A. Gullberg, and A. Janke. (1998). Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* 47:718–727.

Bailey, W. J., D. H. A. Fitch, D. A. Tagle, J. Czelusniak, J. L. Slightom, and M. Goodman. (1991). Molecular evolution of the $\psi\eta$-globin gene locus: Gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* 8:155–184.

Barker, P., I. Dalziel, and B. Storey. (1991). Tectonic development of the Scotia Arc region. Pages 215–248 *in* The geology of Antarctica. (R. Tinger, ed.). Clarendon Press, Oxford.

Bauer, K., and A. Schreiber. (1997). Double invasion of Tertiary Island South America by ancestral NewWorld monkeys? *Biol. J. Linn. Soc*. 60:1–20.

Beard, K. C., T. Qi, M. R. Dawson, B. Wang, and C. Li. (1994). A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. *Nature* 368:604–609.

Bossuyt, F., and M. C. Milinkovitch. (2001). Amphibians as indicators of early tertiary "out-of-India" dispersal of vertebrates. *Science* 292:93– 95.

Bromham, L., and D. Penny. (2003). The modern molecular clock. *Nat. Rev. Genet.* 4:216–224.

Bryant, J. D., and M. C. McKenna. (1995). Cranial anatomy and phylogenetic position of *Tsaganomys altaicus* (Mammalia: Rodentia) from the Hsanda Gol Formation (Oligocene), Mongolia. *Am. Mus. Novitates* 3156:1–42.

Catzeflis, F. M., C.H¨anni, P. Sourrouille, and E. Douzery. (1995). Molecular systematics of hystricognath rodents: The contribution of sciurognath mitochondrial 12S rRNA sequences. *Mol. Phylogenet. Evol.* 4:357–360.

de Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends Ecol. Evol.* 20:68–73.

DeBry, R. W. (2003). Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: The phylogeny of Rodentia (Mammalia). *Syst. Biol.* 52:604–617.

Delsuc, F., S. F. Vizcaino, and E. J. P. Douzery. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: A relaxed molecular clock study within xenarthrans. *BMC Evol. Biol.* 4:11.

Douady, C. J., and E. J. P. Douzery. (2003). Molecular estimation of eulipotyphlan divergence times and the evolution of "Insectivora*." Mol. Phylogenet. Evol.* 28:285–296.

Douzery, E. J. P., F. Delsuc, M.J. Stanhope, and D. Huchon. (2003). Local molecular clocks in

three nuclear genes: Divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J. Mol. Evol.* 57:S201–S213.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.

Fleagle, J. G. (1999). Primate adaptation and evolution, 2nd edition. Academic Press, San Diego.

Flynn, L. J., L.L. Jacobs, and I. U. Cheema. (1986). Baluchimyinae, a new ctenodactyloid rodent subfamily from the Miocene of Baluchistan. *Am. Mus. Novitates* 2841:1–58.

Flynn, J.J., and A.R. Wyss. (1998). Recent advances in South American mammalian paleontology. *Trends Ecol. Evol.* 13:449–454.

Garland, T. J., A. W. Dickerman, C. M. Janis, and J. A. Jones. (1993). Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42:265–292.

Gatesy, J., and M. A. O'Leary. (2001). Deciphering whale origins with molecules and fossils. *Trends Ecol. Evol.* 16:562–570.

Gebo, D. L., M. Dagosto, K. C. Beard, T. Qi, and J. Wang. (2000). The oldest known anthropoid postcranial fossils and the early evolution of higher primates. *Nature* 404:276–278.

Gheerbrant, E., J. Sudre, M. Larochene, and A. Moumni. (2001). First ascertained African "Condylarth" mammals (Primitive ungulates: cf. Bulbulodentata and cf. Phenacodonta) from the earliest Ypresian of the Ouled Abdoun Basin, Morocco. *J. Vert. Paleont.* 21:107–118.

Gingerich, P. D. (1986). Plesiadapis and the delineation of the order Primates. Pages 32–46 *in* Major topics in primate and human evolution (B. Wood, L. Martin, and P. Andrews, eds.). Cambridge University, Cambridge.

Gingerich, P. D., and M. D. Uhen. (1994). Time of origin of primates. *J. Hum. Evol.* 27:443–445.

Glazko, G. V., and M. Nei. (2003). Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20:424–434.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. (1998). Toward a phylogenetic classification of primates based onDNAevidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9:585–598.

Graur, D., and W. Martin. (2004). Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20:80–86.

Groombridge, J. J., C. J. Jones, M. K. Bayes, A. J. van Zyl, J. Carrillo, R. A. Nichols, and M.W. Bruford. (2002).Amolecular phylogeny of African kestrels with reference to divergence across the Indian Ocean. *Mol. Phylogenet. Evol.* 25:267–277.

Gunnell, G. F., and E. R. Miller. (2001). Origin of anthropoidea: Dental evidence and recognition of early anthropoids in the fossil record, with comments on the Asian anthropoid radiation. *Am. J. Phys. Anthropol.* 114:177–191.

Harada, M. L., H. Schneider,M. P. Schneider, I. Sampaio, J. Czelusniak, and M. Goodman. (1995). DNA evidence on the phylogenetic systematics of NewWorld monkeys: Support for the sister-grouping of *Cebus* and *Saimiri* fromtwo unlinked nuclear genes. *Mol. Phylogenet. Evol.* 4:331–349.

Hartenberger, J.-L. (1994). The evolution of the Gliroidea. Pages 19–33 *in* Rodent and

Lagomorph families of Asian origins and diversification (Y. Tomida, C.-K. Li, and T. Setoguchi, eds.). National Science Museum Monograph, Tokyo.

Hasegawa, M., J. L. Thorne, and H. Kishino. (2003). Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. Genes *Genet. Syst.* 78:267–283.

Hassanin, A., and E.J. Douzery. (2003). Molecular and morphological phylogenies of ruminantia and the alternative position of the Moschidae. *Syst. Biol.* 52:206–228.

Hoffstetter, R. (1969). Un primate de l'oligocene inferieur sud-americain: *Branisella boliviana*. *C. R. Hebd. Séances Acad. Sci. Sér.* D 269:434–437.

Hoffstetter, R. (1972). Relationships, origins, and history of the ceboid monkeys and caviomorph rodents: Amodern reinterpretation. *Evol. Biol.* 6:322–347.

Horovitz, I., and A. Meyer. (1995). Systematics of New World monkeys (Platyrrhini, primates) based on 16S mitochondrial DNA sequences: A comparative analysis of different weighting methods in cladistic analysis. *Mol. Phylogenet. Evol.* 4:448–456.

Houle, A. (1999). The origin of platyrrhines: An evaluation of the Antarctic scenario and the floating island model. *Am. J. Phys. Anthropol.* 109:541–559.

Huchon, D., F. Catzeflis, and E. J. P. Douzery. (2000). Variance of molecular datings, evolution of rodents, and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. *Proc. R. Soc. Lond. B* 267:393–402.

Huchon, D., and E. J. P. Douzery. (2001). From the OldWorld to the New World: A molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Mol. Phylogenet. Evol.* 20:238–251.

Huchon, D., O. Madsen, M. J. Sibbald, K. Ament, M. J. Stanhope, F. Catzeflis, W. W. de Jong, and E.J. P. Douzery. (2002). Rodent phylogeny and a timescale for the evolution of Glires: Evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* 19:1053–1065.

Hussain, S. T., H. de Bruijn, and J. M. Leinders. (1978). Middle Eocene rodents from the Kala Chitta Range (Punjab, Pakistan) (III). *Proc. Kon. Ned. Akad. Wetensch. Ser. B* 81:101–112.

Jaeger, J., T. Thein, M. Benammi, Y. Chaimanee, A. N. Soe, T. Lwin, T. Tun, S. Wai, and S. Ducrocq. (1999). A new primate from the Middle Eocene of Myanmar and the Asian early origin of anthropoids. *Science* 286:528–530.

Kay,R. F., C. Ross, and B. A.Williams. (1997). Anthropoid origins. *Science* 275:797–804.

Kishino, H., J. L. Thorne, and W. J. Bruno. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* 18:352–361.

Kumar, S., and S. B. Hedges. (1998).Amolecular timescale for vertebrate evolution. *Nature* 392:917–920.

Lavocat, R. (1969). La systématique des rongeurs hystricomorphes et la dérive des continents. *C. R. Acad. Sci. Paris Sér. D* 269:1496–1497.

Lawver, L. A., and L. M. Gahagan. (2003). Evolution of Cenozoic seaways in the circum-Antarctic region. *Palaeogeo. Palaeoclimatol. Palaeoecol.* 198:11–37.

Lawver, L. A., L. M. Gahagan, and M.F. Coffin. (1992). The development of paleoseaways around Antarctica. The Antarctic paleoenvironment: A perspective on global change. *Am. Geophys.*

Union Antarctic Res. Ser. 56:7–30.

Liu, J.-C., K.D. Makova, R.M. Adkins, S. Gibson, and W.-H. Li. (2001). Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. *Mol. Biol. Evol.* 18:945–953.

Madsen, O., M. Scally, C. J. Douady, D. J. Kao, R.W. DeBry, R. Adkins, H. M. Amrine, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610–614.

Marivaux, L., Y. Chaimanee, S. Ducrocq, B. Marandat, J. Sudre, A. N. Soe, S. T. Tun, W. Htoon, and J. J. Jaeger. (2003). The anthropoid status of a primate from the Late Middle Eocene Pondaung Formation (Central Myanmar): Tarsal evidence. *Proc. Natl. Acad. Sci. USA* 100:13173–13178.

Marivaux, L., M. Vianey-Liaud, J. -L.Welcomme, and J. -J. Jaeger. (2002). The role of Asia in the origin and diversification of hystricognathous rodents. *Zool. Scripta* 31:225–239.

Martin, R. D. (1993). Primate origins: Plugging the gaps. *Nature* 363:223–234.

Martin, T. (1994). African origin of caviomorph rodents is indicated by incisor enamel microstructure. *Paleobiology* 20:5–13.

McKenna, M. C., and S. K. Bell. (1997). Classification of mammals above the species level. Columbia University Press, New York.

Montgelard, C., C. A. Matthee, and T. J. Robinson. (2003). Molecular systematics of dormice (Rodentia: Gliridae) and the radiation of *Graphiurus* in Africa. *Proc. R. Soc. Lond. B Biol. Sci.* 270:1947–1955.

Mouchaty, S. K., F. Catzeflis, A. Janke, and U. Arnason. (2001). Molecular evidence of an African Phiomorpha–South American Caviomorpha clade and support for Hystricognathi based on the complete mitochondrial genome of the cane rat (*Thryonomys swinderianus*). *Mol. Phylogenet. Evol.* 18:127–135.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.

Nagy, Z. T., U. Joger, M. Wink, F. Glaw, and M. Vences. (2003). Multiple colonization of Madagascar and Socotra by colubrid snakes: Evidence from nuclear and mitochondrial gene phylogenies. *Proc. R. Soc. Lond. B Biol. Sci.* 270:2613–2621.

Nedbal, M. A., M.W. Allard, and R. L. Honeycutt. (1994). Molecular systematics of hystricognath rodents: Evidence from the mitochondrial 12S rRNA gene. *Mol. Phylogenet. Evol.* 3:206–220.

Nei, M., and G. V. Glazko. (2002). Estimation of divergence times for a few mammalian and several primate species. *J. Hered.* 93:157–164.

Nikaido, M., K. Kawai, Y. Cao, M. Harada, S. Tomita, N. Okada, and M. Hasegawa. (2001). Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53:508–516.

Philippe, H. (1993). MUST: A computer package of management utilities for sequences and

trees. *Nucleic. Acids Res.* 21:5264–5272.

Posada, D., and K. A. Crandall. (1998). ModelTest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.

Poux, C., and E. J. P. Douzery. (2004). Primate phylogeny, evolutionary rate variations, and divergence times: A contribution from the nuclear gene IRBP. *Am. J. Phys. Anthropol.* 124:1–16.

Poux, C., O. Madsen, E. Marquard, D. R. Vieites, W. W. de Jong, and M. Vences. (2005). Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores and rodents as inferred from nuclear genes. *Syst. Biol.* 54:719–730.

Ronquist, F., and J. P Huelsenbeck. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572– 1574.

Roos, C., J. Schmitz, and H. Zischler. (2004). Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl. Acad. Sci. USA* 101:10650– 10654.

Rosenberger, A. L., W. C. Hartwig, and R. G. Wolff. (1991). *Szalatavus attricuspis*, an early platyrrhine primate. *Folia Primatol.* 56:225–233.

Ross, C. F. (2000). Into the light: The origin of Anthropoidea. *Annu. Rev. Anthropol.* 29:147–194.

Sanmartín, I., and F. Ronquist. (2004). Southern hemisphere biogeography inferred by event-based models: Plant versus animal patterns. *Syst. Biol.* 53: 216–243.

Schneider, H. (2000). The current status of the NewWorld monkey phylogeny. *An. Acad. Bras. Cienc.* 72:165–172.

Schneider, H., F. C. Canavez, I. Sampaio, M. A. Moreira, C. H. Tagliaro, and H. N. Seuanez. (2001). Can molecular data place each neotropical monkey in its own branch? *Chromosoma* 109:515–523.

Schneider, H., I. Sampaio, M. L. Harada, C. M. Barroso, M. P. Schneider, J. Czelusniak, and M. Goodman. (1996). Molecular phylogeny of the NewWorld monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences. *Am. J. Phys. Anthropol.* 100:153–179.

Schrago, C. G., and C. A. Russo. (2003). Timing the origin of new world monkeys. *Mol. Biol. Evol.* 20:1620–1625.

Seiffert, E. R., E. L. Simons, and C. V. M. Simons. (2003). Phylogenetic, biogeographic, and adaptative implications of new fossil evidence bearing on crown anthropoid origins and early stem catarrhine evolution. Pages 157–181 *in* Anthropoid origins (C. F. Ross, and R. F. Kay, eds.). Kluwer Academic/Plenum Publishers, New York.

Shoshani, J., C. P. Groves, E. L. Simons, and G. F. Gunnell. (1996). Primate phylogeny: Morphological vs. molecular results. *Mol. Phylogenet. Evol.* 5:102–154.

Simons, E. L. (1997). Preliminary description of the cranium of *Proteopithecus sylviae*, an Egyptian late Eocene anthropoidean primate. *Proc. Natl. Acad. Sci. USA* 94:14970–14975.

Smith, A. G., D. G. Smith, and B. M. Funnel. (1994). Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.

Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. (2002). Rate

heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. USA* 99:4430–4435.

Springer,M. S., R.W. DeBry, C. Douady, H.M. Amrine, O. Madsen,W. W. de Jong, and M. J. Stanhope. (2001). Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18:132–143.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100:1056–1061.

Springer,M. S., M. Scally, O. Madsen,W.W. de Jong, C. J. Douady, and M. J. Stanhope. (2004a). The use of composite taxa in supermatrices. *Mol. Phylogenet. Evol.* 30:883–884.

Springer, M. S., M. J. Stanhope, O. Madsen, and W. W. de Jong. (2004b). Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* 19:430–438.

Steiper, M. E., N. M. Young, and T. Y. Sukarna. (2004). Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl. Acad. Sci. USA* 101:17021–17026.

Swofford, D. L. (1999). PAUP∗: Phylogenetic analysis using parsimony (∗and other methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

Takahata, N., and Y. Satta. (1997). Evolution of the primate lineage leading to modern humans: Phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* 94:4811–4815.

Takai, M., and F. Anaya. (1996). New specimens of the oldest fossil platyrrhine, *Branisella boliviana*, from Salla, Bolivia. *Am. J. Phys. Anthropol.* 99:301–317.

Takai, M., F. Anaya, N. Shigehara, and T. Setoguchi. (2000). New fossil materials of the earliest NewWorld monkey, *Branisella boliviana*, and the problem of platyrrhine origins. *Am. J. Phys. Anthropol.* 111:263–281.

Tavare, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. (2002). Using the fossil recordto estimate the age of the lastcommonancestor of extant primates. *Nature* 416:726–729.

Thorne, J. L., and H. Kishino. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–1657.

Tullberg, T. (1899). Ueber das System der Nagetiere: Eine phylogenetische Studie. *Nova Acta Reg. Soc. Sci. Upsala Ser.* 3:1–514

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer. (2003). Multiple overseas dispersal in amphibians. *Proc. R. Soc. Lond. B Biol. Sci.* 270:2435–2442.

von Dornum, M., and M. Ruvolo. (1999). Phylogenetic relationships of the New World monkeys (primates, Platyrrhini) based on nuclear G6PD DNA sequences. *Mol. Phylogenet. Evol.* 11:459–476.

Vucetich, M. G., D. H. Verzi, and J.-L. Hartenberger. (1999). Review and analysis of the radiation of the South American Hystricognathi (Mammalia, Rodentia). *C. R. Acad. Sci. Paris Earth Planet. Sci.* 329:763–769.

Wyss, A. R., J. J. Flynn, M. A. Norell, C. C. Swisher III, R. Charrier,M. J. Novacek, and M. C. McKenna. (1993). South America's earliest rodent and recognition of a new interval of mammalian evolution. *Nature* 365:434–437.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11:367–372.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555–556.

Yang, Z., and A. D. Yoder. (2003). Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* 52:705–716.

Yoder, A. D., M. M. Burns, S. Zehr, T. Delefosse, G. Veron, S. M. Goodman, and J. J. Flynn. (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421:734–737.

Yoder, A. D., M. Cartmill, M. Ruvolo, K. Smith, and R. Vilgalys. (1996). Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. USA.* 93:5122–5126.

Yoder, A. D., and Z. Yang. (2004). Divergence dates for Malagasy lemurs estimated from multiple gene loci: Geological and evolutionary context. *Mol. Ecol.* 13:757–773.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–693.

Chapter 5

Asynchronous Colonization of Madagascar by the Four Endemic Clades of Primates,Tenrecs, Carnivores, and Rodents as Inferred from Nuclear Genes

Céline Poux[1], Ole Madsen[1], Elisabeth Marquard[1,2], David R. Vieites[3], Wilfried W. de Jong[1,2], and Miguel Vences[2]

[1] Department of Biochemistry 161, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

[2] Institute for Biodiversity and Ecosystem Dynamics, Zoological Museum, University of Amsterdam, 1092 AD Amsterdam, The Netherlands.

[3] Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, California 94720-3160, USA.

Systematic Biology 54:719–730. 2005

ABSTRACT

Madagascar harbors four large adaptive radiations of endemic terrestrial mammals: lemurs, tenrecs, carnivorans, and rodents. These rank among the most spectacular examples of evolutionary diversification, but their monophyly and origins are debated. The lack of Tertiary fossils from Madagascar leaves molecular studies as most promising to solve these controversies. We provide a simultaneous reconstruction of phylogeny and age of the four radiations based on a 3.5-kb data set from three nuclear genes (ADRA2B, vWF, and AR). The analysis supports each as a monophyletic clade, sister to African taxa, and thereby identifies four events of colonization out of Africa. To infer the time windows for colonization, we take into account both the divergence from the closest noninsular sister group and the initial intrainsular radiation, which is a novel but conservative approach in studies of the colonization history of Madagascar. We estimate that lemurs colonized Madagascar between 60 million years ago (Mya) (split from lorises) and 50 Mya (lemur radiation) (70–41 Mya taking 95% credibility intervals into account), tenrecs between 42 and 25 Mya (50–20 Mya), carnivorans between 26 and 19 Mya (33–14 Mya), and rodents between 24 and 20 Mya (30–15 Mya). These datings suggest at least two asynchronous colonization events: by lemurs in the Late Cretaceous–Middle Eocene, and by carnivorans and rodents in the Early Oligocene–Early Miocene. The colonization by tenrecs may have taken place simultaneously with either of these two events, or in a third event in the Late Eocene–Oligocene. Colonization by at least lemurs, rodents, and carnivorans appears to have occurred by overseas rafting rather than via a land bridge hypothesized to have existed between 45 and 26 Mya, but the second scenario cannot be ruled out if credibility intervals are taken into account.

INTRODUCTION

The study of adaptive radiations on islands has been essential for understanding processes of evolutionary diversification (Grant 1998; Losos et al. 1998). Reconstructing the origin and phylogeny of endemic island taxa provides crucial insight into transoceanic dispersal mechanisms and in the factors triggering radiation processes. Among major islands, Madagascar has long been renowned for

the uniqueness of its fauna and flora (Myers et al. 2000), with a species-level endemism in non-flying vertebrates of over 95% that is mainly due to a few speciose endemic radiations (e.g. Bossuyt & Milinkovitch 2001; Nagy et al. 2003; Vences et al. 2003). Madagascar became isolated from India 96 to 84 million years ago (Mya), and overland connections with Africa were severed approximately 160 to 158 Mya (Briggs 2003).

Terrestrial mammals are represented in Madagascar by about 100 endemic species (Goodman et al. 2003) belonging to four taxonomic groups: lemurs, tenrecs, nesomyine rodents, and carnivorans (Fig.1). These represent four of the 16 orders of land-dwelling placental mammals. Recent molecular studies have provided compelling evidence that Malagasy lemurs and carnivorans, despite their striking morphological diversity, are two monophyletic groups that presumably originated from single African ancestors (Yoder et al. 2003; Roos et al. 2004). However, morphological and molecular data are inconsistent with regard to the monophyly and intercontinental relationships of Malagasy tenrecs (Eisenberg 1981; Asher 1999; Douady & Douzery 2003; Olson & Goodman 2003) and nesomyine rodents (Lavocat 1978; Dubois et al. 1998; Jansa et al. 1999; Michaux et al. 2001; Jansa & Weksler 2004; Steppan et al. 2004), possibly because of extraordinary similarities to non-Malagasy forms. The Malagasy tenrec lineage has spawned hedgehog-like tenrecines, mole- and shrew-like oryzoryctines, and a semiaquatic form (*Limnogale*), whereas nesomyine rodents comprise vole- and gerbil-like species (*Brachyuromys* and *Macrotarsomys*) as well as arboreal and giant jumping rats (*Brachytarsomys* and *Hypogeomys*).

Fossil evidence to help resolve the origin of Madagascar's mammals is scarce. Relevant fossils are absent from Madagascar for the whole of the Tertiary period, and the rich findings from the Late Cretaceous include gondwanatheres, multituberculates, and marsupials, but no fossils related to extant taxa (Krause et al. 1997a, 1997b; Krause 2001). The extant mammal groups probably arrived during the Cenozoic after the complete isolation of Madagascar (Krause et al. 1997a). However, most terrestrial mammals are poor over-water dispersers as indicated by their rareness on isolated oceanic islands (Lawlor 1986). To reconcile these facts, a land bridge has been proposed that might have connected Africa and Madagascar from ~45 to ~26 Mya (McCall 1997). Alternatively, mammals may have reached Madagascar by "rafting" or island-hopping (e.g. Krause et al. 1997a).

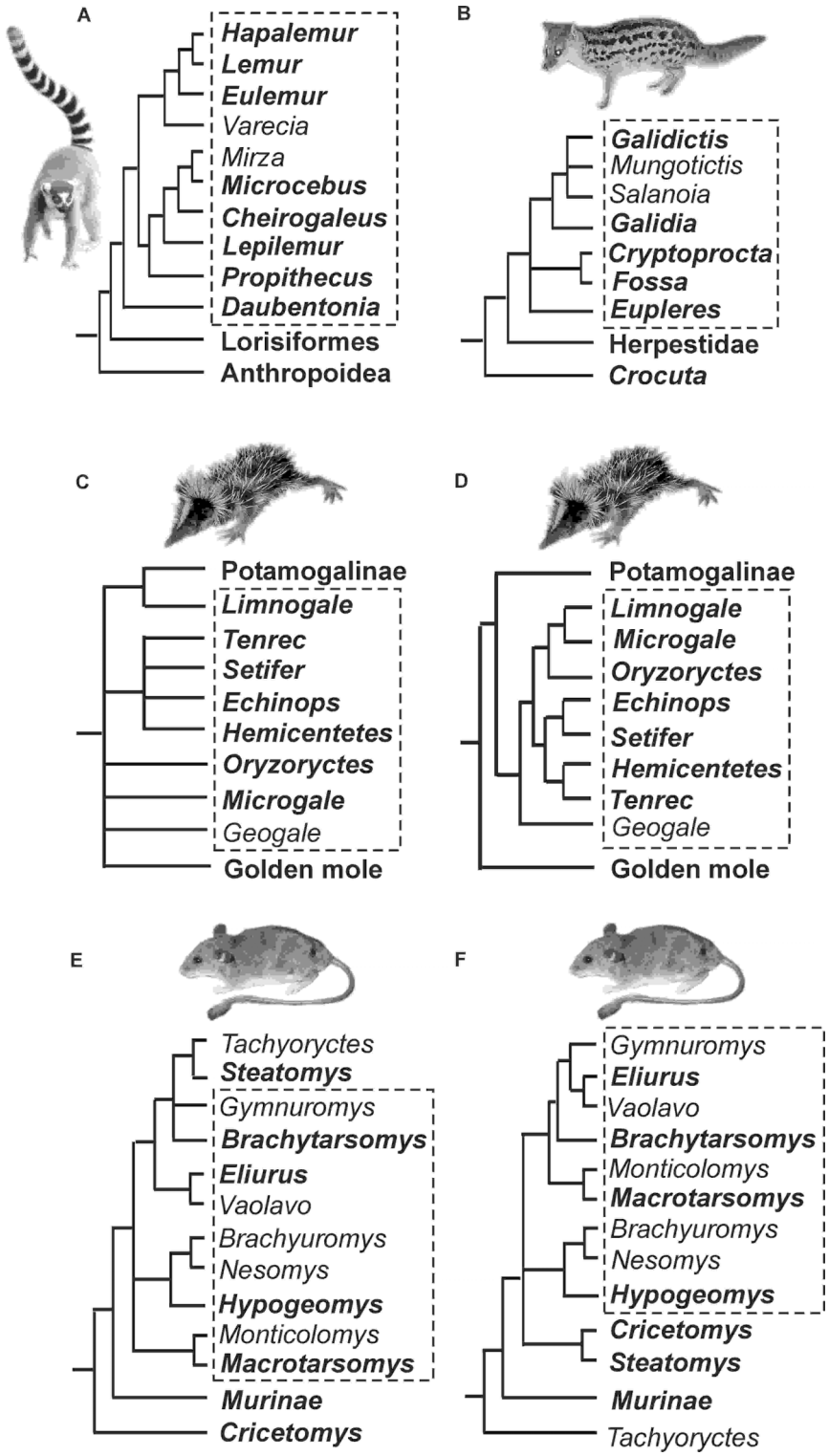

Figure 1: Phylogenetic hypotheses for lemurs (A), carnivores (B), tenrecs (C, D) and rodents (E, F) of Madagascar. Dashed boxes enclose endemic Malagasy taxa. Taxa in bold were included in the present study. Molecular data unequivocally suggested a monophyletic origin for lemurs (A) and Malagasy carnivorans (B) (Yoder et al. 1996, 2003). Morphological data indicated paraphyly of Malagasy tenrecs (C; consensus of alternative morphological trees; Asher 1999), but molecular data support their monophyly (D; Olson & Goodman 2003). Analysis of cytochrome *b* sequences suggested paraphyly of Malagasy rodents (E; Jansa et al. 1999), whereas IRBP sequences could not resolve their relationships (F; Jansa & Weksler 2004). In the case of tenrecs and rodents some taxa are excluded to make trees comparable. Photos show *Lemur catta* (A), *Fossa fossana* (B), *Hemicentetes semispinosus* (C, D), and *Eliurus* sp. (E, F).

We here apply a DNA sequence data set of almost 3.5 kb from three independent nuclear genes to the reconstruction of phylogeny and age of the four Malagasy mammalian radiations and find compelling support for their respective monophyly. We argue in favor of a more conservative approach to date ages of island colonization by taking into account both the divergence from the closest noninsular sister group and the deepest intrainsular divergence, and apply this method to test alternative hypotheses for the origin of endemic Malagasy mammals. Because all four Malagasy clades are included in the same analyses, the obtained molecular datings are directly comparable and strongly support independent colonizations by overseas rafting.

MATERIALS AND METHODS

Sampling, DNA amplification and sequencing

Fragments of the intronless gene of alpha 2B adrenergic receptor (ADRA2B), of exon 28 of the vonWillebrand factor (vWF) gene, and of exon 1 of the androgen receptor (AR) gene were amplified and sequenced. These genes were selected because (i) they are located in the nuclear genome, as single-copy genes in at least human and mouse, and such genes are generally superior to mitochondrial genes for reconstruction of ancient relationships (Springer et al. 2001) and for time estimations (Glazko & Nei 2003); (ii) a considerable number of sequences are already available for ADRA2B and vWF and have been useful in deeper mammalian phylogeny; and (iii) they are functionally and genetically unrelated. We selected 60 mammal species to represent for each of the three genes (i) the major lineages of all four Malagasy mammal radiations; (ii) their potential sister groups; (iii) groups needed for multiple calibration of the molecular clock; (iv) other basal mammal clades; and (v) appropriate outgroups. A total of 103 new sequences were obtained and complemented with 72 sequences from GenBank (Appendix 1). The full data matrix is available from Treebase (accession number: M2279).

Genomic DNA was isolated from ethanol-preserved tissue, following the protocols of either the DNeasy Tissue Kit (Qiagen) or the Wizard SV Genomic DNA Purification System (Promega). Fragments of the intronless ADRA2B gene and of exon 28 of the vWF gene were amplified using previously published primers (Porter

et al. 1996; Springer et al. 1997). Two new vWF primers were designed for some species (vWF-for and vWFrev), and exon 1 of the AR gene was amplified with the primers F-AR1 and R-AR1 (Appendix 2, available at www.systematicbiology.org). Polymerase chain reactions (PCRs) were performed on 50 to 200 ng DNA with Expand DNA polymerase (Expand High Fidelity PCR system, Roche) using the following program: 2 min at 94◦C; 30 to 35 cycles of 15 s at 94◦C, 1 min at 52 to 58◦C, 56–61◦C, or 55–59◦C (for vWF, A2AB, or AR, respectively), and 1 min 30 s at 72◦C; and a final step of 2–10 min at 72◦C. DMSO (1.3% to 2.5%) and/or betaine (1 M) was added for some samples. PCR products were purified from a 1% agarose gel, using GFX PCR DNA & Gel Band Purification Kit (Amersham Biosciences) and reamplified if necessary. Gel-extracted PCR products were sequenced directly on ABI 3700 or 3730 96-capillary sequencers (Applied Biosystems).

Some specimens were polymorphic for glutamine tracks in AR or a glutamic acid track in ADRA2B. These PCR-products were cloned into a pGEM-T Vector (Promega), transformed into competent *E. coli* TOP 10 cells, and positive clones sequenced. Internal primers were used to get complete sequences of both strands. For one tenrec, *Microgale*, no vWF sequence could be obtained; this species was excluded from the molecular dating.

Phylogenetic analyses

Sequences were assembled with PreGAP and GAP4 (Staden package, http://www.mrc-lmb.cam.ac.uk/pubseq). Alignments were obtained using GCG PILEUP (Wisconsin Package Version 10.3, Accelrys Inc.) and manually adjusted considering amino acid properties. Amino acids repeats and sites not sequenced or gapped in more than 25% of the taxa were excluded from analysis. This resulted in a data set of 1134 bp for ADRA2B, 1141 bp for vWF, and 1212 bp for AR.

Phylogenetic reconstructions on the concatenated data set were performed by maximum likelihood (ML) with PAUP*, version 4b10 (Swofford 2003), and Bayesian analyses with MRBAYES, version 3.0b4 (Huelsenbeck & Ronquist 2001). The best fitting model under theML criterion was selected by the hLRT output of ModelTest, version 3.5 (Posada & Crandall 1998). ML analyses included heuristic searches with a neighbor-joining starting tree and tree bisection-reconnection branch swapping. Node stability was estimated by 100 nonparametric bootstrap replicates

(Felsenstein 1985).

A major advantage of Bayesian phylogenetic inference is the possibility of partitioning the data, giving each partition its own best-fitting model of sequence evolution. However, overpartitioning may introduce unnecessary sampling variances, which could influence the phylogenetic estimates. For the nine possible codon partitions (each codon position of each gene), ModelTest was used to calculate the best fitting model of sequence evolution. As further explained in Table 1, codon partitions with similar models and model parameters were merged, resulting in six partitions for the Bayesian analyses: the first codon positions of ADRA2B and vWF, and the second positions of AR; the second positions of ADRA2B and vWF; the first position of AR; and the third codon positions of each gene separately. Four Markov chains were run simultaneously for 1,000,000 and 500,000 generations, with initial equal probabilities for all trees and starting with a random tree. Tree sampling frequency was each 20 generations and the consensus tree with posterior probabilities was calculated after removal of the first 2500 trees ("burn in" as determined from the likelihood values).

Table 1. Best-fitting evolutionary model for each codon position. Best models and parameters were found with the hierarchical likelihood ratio test as implemented in ModelTest 3.5 for each codon position of the three gene fragments. Codon positions with similar model and model parameters were regrouped into the same partition and resulted in six partitions: (1) first codon positions of ADRA2B and vWF, and second positions of AR; (2) second positions of ADRA2B and vWF; (3) first positions of AR; and (4–6) third codon positions of each gene separately. Codon positions were merged into the same partition when none of their model parameters (e.g. TRatio of position 1 compared to TRatio of position 2, PInvar 1 to PInvar 2, etc.) differed by more that 100%. The maximum difference between model parameters within one partition was 58%. TRatio, transition/transversion ratio; rmat, rate matrix; π, frequency of base; PInvar, proportion of invariable sites; alpha, shape of gamma distribution.

Gene	Codon position	Length	π_A	π_C	π_G	Best model	TRatio or Rmat	alpha	PInvar	Partition Number
ADRA2B	1	378	0.25	0.25	0.25	K80+Γ	1.47	0.42	0	1
	2	378	0.19	0.32	0.23	TrN+Γ+I	(1.0 4.4 1.0 1.0 2.9)	1.08	0.45	2
	3	378	0.13	0.38	0.32	TVM+Γ	(1.0 4.5 1.7 0.4 4.5)	2.46	0	4
vWF	1	381	0.27	0.30	0.31	HKY+Γ	1.30	0.59	0	1
	2	380	0.30	0.28	0.16	TrN+Γ+I	(1.0 5.8 1.0 1.0 4.6)	0.81	0.31	2
	3	380	0.10	0.38	0.38	TVM+Γ	(1.6 8.1 3.5 0.6 8.1)	2.9	0	5
AR	1	404	0.23	0.26	0.30	TrN+Γ	(1.0 5.3 1.0 1.0 3.7)	0.62	0	3
	2	404	0.25	0.34	0.21	HKY+Γ	1.1	0.55	0	1
	3	404	0.24	0.28	0.24	HKY+Γ	2.5	1.65	0	6

Molecular dating

We used the Bayesian approach (Thorne et al. 1998) as implemented in the MULTIDIVTIME program package (Thorne & Kishino 2002), which relaxes the molecular clock by allowing continuous autocorrelation of substitution rates among the branches of the phylogenetic tree. This approach estimates rates accurately (Ho et al. 2005), and was here chosen instead of penalized likelihood (Sanderson 2002) because the MULTIDIVTIME software does not require the root of the tree to be fixed at a particular date but estimates its age starting from a prior value.

The concatenated sequence data set was partitioned into the same six categories as for the Bayesian phylogenetic analyses and branch lengths calculated under the F84+gamma model of sequence evolution, which is the most complex model available in MULTIDIVTIME. The prior for the root was set at 100 Mya. Markov chain Monte Carlo (MCMC) analyses were run for 3,000,000 and 1,000,000 generations with a "burn in" of 100,000 generations. The chains were sampled every 100 generations. To assess the influence of our particular partitioning on the dating results, we performed additional analyses using five other partitioning schemes, and without partitioning, running MCMC analyses for 1,000,000 generations. The results of these six supplementary analyses were close to each other. Notably, all datings for the nodes that we were interested in remained within the 95% credibility intervals of the datings obtained in the original analysis using six partitions (cf. Table 2). Our conclusions are therefore not affected by the choice of our partitioning.

Six well-established fossil constraints on divergence times were used: (i) a minimum of 54 and a maximum of 65 Mya for the base of Paenungulata (Gheerbrant et al. 2001); (ii) a minimum of 37 Mya for the split between ochotonids and leporids (McKenna & Bell 1997); (iii) a minimum of 63 and a maximum of 90 Mya for the radiation of primates (Martin 1993; Gingerich & Uhen 1994; Tavare et al. 2002); (iv) a minimum of 50 and a maximum of 63 Mya for the split between feliform and caniform carnivorans (Benton 1993; McKenna & Bell 1997); (v) a minimum of 54 and a maximum of 58 Mya for the split beween hippomorph and ceratomorph Perissodactyla (Garland 1993); (vi) a minimum of 55 and a maximum of 65 Mya for the base of Cetartiodactyla (Gatesy & O'Leary 2001). To assess the reciprocal compatibility of these calibrations, calculations were repeated after their removal one by one, the Markov chains being sampled 1,000,000 times.

RESULTS AND DISCUSSION

Assessing relationships of Malagasy mammals

To determine the phylogenetic relations of Malagasy mammals we analyzed sequences from the nuclear genes for ADRA2B, vWF, and AR. Our sampling included 13 of the 18 orders of placental mammals, and two marsupial outgroup orders (Appendix 1). Phylogenetic analysis of the concatenated 3487-bp data set, by maximum likelihood (ML) and Bayesian methods, recovered inter- and intraordinal relationships (Fig.2) in perfect agreement with more comprehensive recent phylogenies (reviewed by Springer et al. 2004). These include the superordinal clades Afrotheria, Boreoeutheria, Euarchontoglires, and Laurasiatheria, as well as Glires (rodents and lagomorphs) and Paenungulata (elephants, sea cows and hyraxes). This concordance with previous results increases the confidence in the phylogenetic relationships newly deduced here. Our analysis found each of the four endemic Malagasy mammal radiations to be monophyletic, with maximal bootstrap percentages and posterior probabilities (BP=100, PP=1.00) for Malagasy tenrecs, rodents, and carnivorans. Only the monophyly of the lemurs was poorly supported (BP=47, PP=0.86), but corroborated by a unique 15-bp deletion in the vWF sequence of all Lemuriformes, including the most basal aye-aye (*Daubentonia*) (Appendix 3A, available at www.systematicbiology.org).

The monophyly of Malagasy carnivorans and their relationship to herpestids, here represented by *Suricata*, confirmed previous molecular data (Yoder et al. 2003). The same applied to the monophyletic lemurs that are sister to the Lorisiformes, here represented by *Nycticebus* (Yoder et al. 1996, 2003). Our data further confirmed the phylogenetic relationships among Malagasy carnivoran and lemuriform taxa (Yoder et al.1996, 2003; Pastorini et al. 2003; Roos et al. 2004), and provide the first compelling evidence for a close relationship of the specialized worm-eating civet *Eupleres* to *Fossa* (BP=100, PP=1.00).

Monophyly of Malagasy tenrecs was strongly supported in our analyses and relations among included taxa were resolved completely and with high support (Fig.2), whereas morphological data have been ambiguous in this respect (Asher 1999). The African otter shrews, here represented by *Micropotamogale*, were found

as sister group of all Malagasy tenrecs. The semiaquatic web-footed tenrec *Limnogale* which morphologically resembles the otter shrew (Asher 1999) actually appeared closely related to the shrew tenrec *Microgale* (BP=100, PP=1.00). This relationship was corroborated by a molecular synapomorphy, a shared 3-bp deletion in the ADRA2B gene (Appendix 3B, available at www.systematicbiology.org).

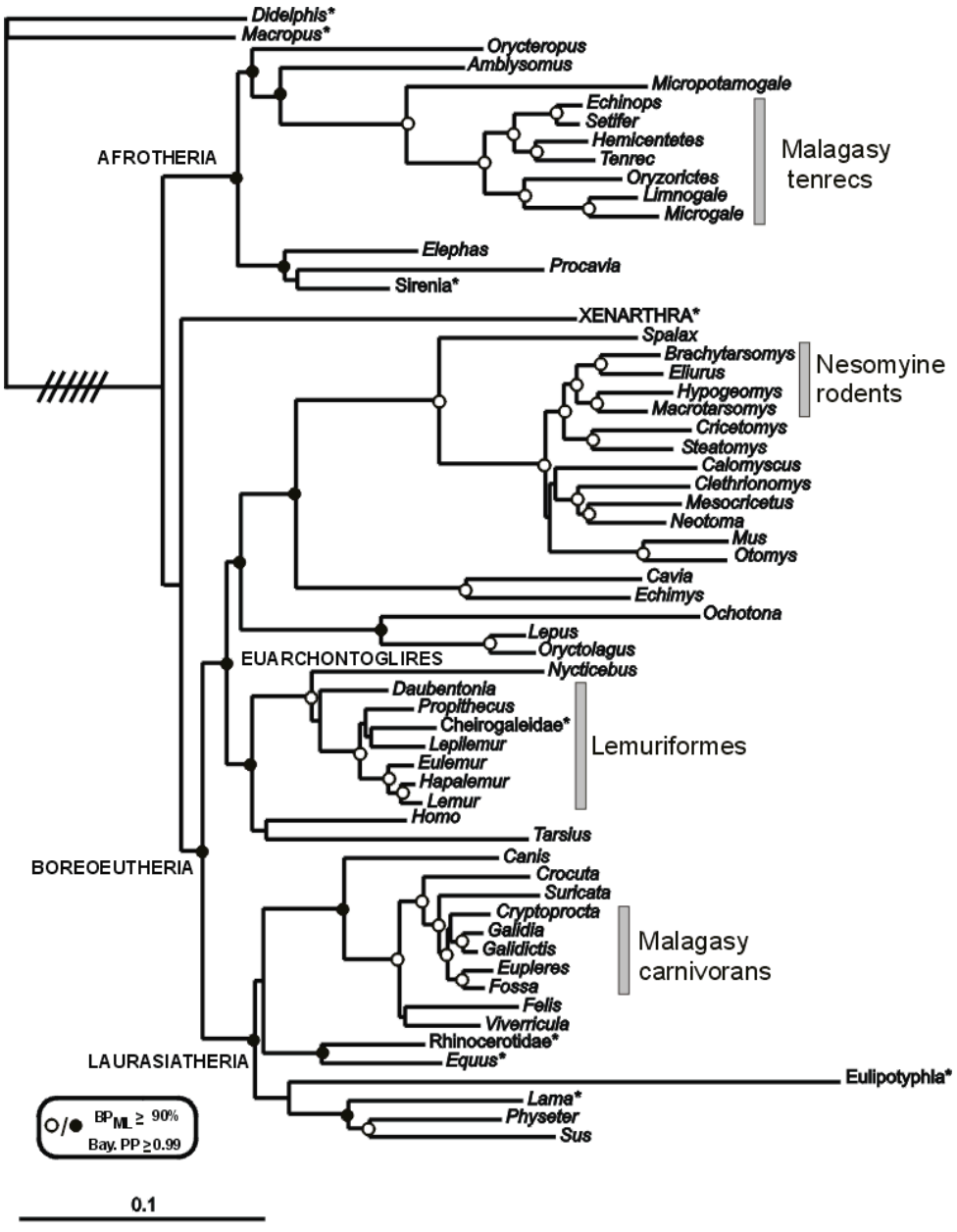

Figure 2: Phylogenetic relationships of Malagasy mammals as inferred by maximum likelihood analysis of the concatenated 3487-bp data set of ADRA2B, vWF, and AR sequences. Bayesian analyses result in an identical topology. Nodes receiving high support (BP ≥ 90% and PP ≥ 0.99) are marked with circles; filled circles correspond with generally accepted ordinal and superordinal relationships. The length of the branch connecting eutherians to the marsupial outgroup was reduced six times. Asterisks mark taxa represented by different species in the concatenated sequences (Appendix 1).

Nesomyine rodents belong to the Muridae, the most speciose family of mammals. Previous studies identified various major clades within nesomyines but have been unable to resolve relationships between these and several non-Malagasy murid taxa (Jansa et al. 1999; Jansa & Weksler 2004). Our analysis included representatives of each of these clades (cf. Fig.1) and the monophyly of Malagasy rodents was firmly established (Fig.2). Their sister group was a clade comprising the African murids *Steatomys* (Dendromurinae) and *Cricetomys* (Cricetomyinae) (BP=100, PP=1.00).

Hence, monophyly and relations to African taxa were unambiguously suggested for Malagasy tenrecs and nesomyine rodents, where the evidence was so far controversial, and confirmed for the Malagasy carnivorans. The sister group of the Lemuriformes contains African an Asian taxa, but an African origin of the strepsirrhine clade is now supported (Seiffert et al. 2003; Roos et al. 2004). This strongly suggests that each of the four clades originated by a single colonization event out of Africa.

Timing the origins and radiations of Malagasy mammals

DNA sequences can be used in various statistical approaches to estimate times of divergence (Hedges & Kumar 2003). Such molecular datings face two main problems. First, the assumption of evolutionary rate constancy is in general not valid (Bromham & Penny 2003), as obvious in our data set from branch lengths in Figure 2. Second, the fossil ages used in the analyses may not be accurate (Graur & Martin 2004). Biases can in both cases result in erroneous time estimates. We here simultaneously used six independent fossil calibrations, specified in Materials and Methods, in a relaxed clock approach that takes into account the variations of the molecular substitution rate. By constraining the time estimates on the fossil calibrations as ranges rather than fixed values, the method takes also the paleontological uncertainties into account.

The age of colonization of Madagascar has usually been seen as equivalent either to the initial diversification of the Malagasy lineages (Yoder et al. 1996, 2003; Roos et al. 2004) or to the split from their non-Malagasy sister group (Nagy et al. 2003; Vences et al. 2003). However, a radiation may take place long after the initial colonization, or early radiations may go extinct. Moreover, the extant mainland sister

group of an insular clade is not necessarily its closest mainland relative which may have gone extinct (Fig.3A). The same rationale has been applied for the colonization of South America by rodents and primates (Poux et al. 2006). Hence, in order to obtain a conservative and more reliable estimate of the time period during which colonization has occurred, we here suggest that the two divergence times for the latest outgroup split and the earliest ingroup split, and their 95% credibility intervals, need to be taken into account (Fig.3B).

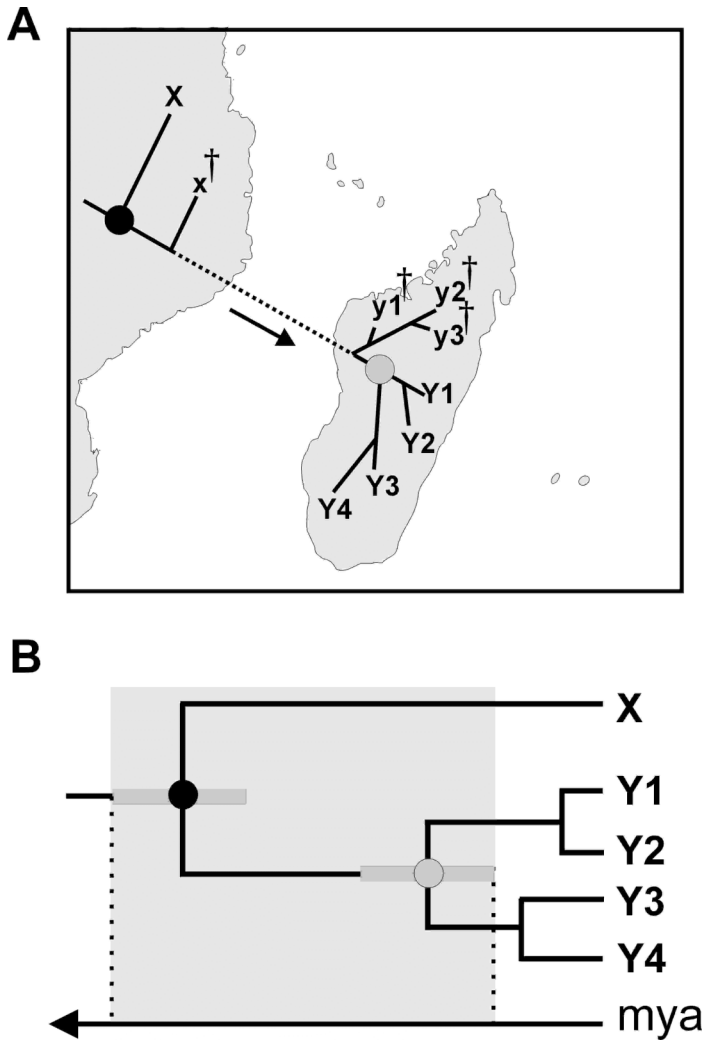

Figure 3: Estimating the colonization time of an island (here Madagascar) using molecular clock data from extant taxa. (A) Molecular datings provide estimates for the time of divergence of the extant insular taxa Y from their nearest extant noninsular sister taxon X (black circle) and for the earliest divergence among the extant insular taxa Y1–Y4 (grey circle). These two estimates provide the maximum time window for possible colonization. Any of the estimates may be close to the actual time of colonization, but extinct species x from the mainland may have been closer to the colonizing ancestor, and early radiations in Madagascar (y1–y3) may have gone extinct. Fossil data might therefore shorten the time window for colonization (dashed line). (B) The most conservative window of possible colonization times (shaded area) is given by the upper 95% confidence interval of the first estimate (black circle) and the lower 95% confidence interval of the second estimate (grey circle).

Table 2. Posterior estimates of divergence ages (± standard deviation) inferred from the concatenated data set using the Bayesian relaxed molecular clock method. The 95% credibility intervals are given between brackets.

Radiation or branching (/)	Calibration time frame (mya)[a]	All calibration points[b]	Removal of the following calibration point during the analysis[c]:					
			1 Paenungu-lata	2 Ochotona / Leporidae split	3 Primates	4 Feliformia / Caniformia split	5 Perisso-dactyla	6 Cetartio-dactyla
Malagasy tenrecs		25.3 ± 3.1 (31.8-19.7)	26.7 ± 3.6	25.3 ± 3.0	25.4 ± 3.1	25.4 ± 3.0	25.3 ±3.0	25.2 ± 3.0
Malagasy tenrecs / Potamogalines		41.8 ± 4.1 (50.3-34.1)	43.9 ± 5.1	41.7 ± 4.1	41.9 ± 4.2	41.8 ± 4.1	41.7 ± 4.1	41.5 ± 4.1
Nesomyine rodents		20.1 ± 2.6 (25.7-15.4)	20.6 ± 2.8	20.2 ± 2.7	20.3 ± 2.6	20.1 ± 2.6	20.0 ± 2.7	19.9 ± 2.7
Nesomyine rodents / (*Cricetomys* + *Steatomys*)		23.5 ± 2.9 (29.6-18.2)	24.1 ± 3.0	23.6 ± 2.9	23.7 ± 2.9	23.5 ± 2.9	23.4 ± 3.0	23.2 ± 2.9
Malagasy lemurs		49.6 ± 4.4 (58.5-41.1)	50.6 ± 4.6	49.6 ± 4.4	49.8 ± 4.5	49.4 ± 4.4	49.3 ± 4.6	49.0 ± 4.6
Malagasy lemurs / Lorisiformes		60.4 ± 4.6 (69.6-51.6)	61.5 ± 4.8	60.4 ± 4.6	60.6 ± 4.8	60.2 ± 4.7	60.0 ± 4.8	59.6 ± 4.9
Malagasy carnivorans		19.0 ± 2.7 (24.8-14.1)	19.1 ± 2.8	19.0 ± 2.7	19.0 ± 2.7	18.9 ± 2.9	18.9 ± 2.7	18.9 ± 2.7
Malagasy carnivorans / *Suricata*		25.9 ± 3.2 (32.5-20.1)	26.1 ± 3.2	25.9 ± 3.2	25.9 ± 3.2	25.7 ± 3.4	25.7 ± 3.2	25.7 ± 3.1
1 Paenungulata	54-65 mya	60.9 ± 2.8	64.8 ± 5.7					
2 Ochotona / Leporidae	> 37 mya	52.8 ± 4.7		52.8 ± 4.7				
3 Primates	63-90 mya	78.9 ± 4.5			79.1 ± 4.7			
4 Feliformia / Caniformia	50-63 mya	55.6 ± 3.1				55.2 ± 4.0		
5 Perissodactyla	54-58 mya	55.9 ± 1.1					54.8 ± 2.3	
6 Cetartiodactyla	55-65 mya	58.6 ± 2.5						56.2 ± 4.3

[a] Paleontological time constraints used as calibrations, numbered as indicated in Figure 4.
[b] Estimates of the age of the Malagasy lineages as used in Figure 4 and in the text.
[c] The influence of each calibration point was tested by computing divergence ages after removing that calibration point. The lower panel shows that all six calibrations points are correctly recovered when the point itself is excluded from the constraints. The recovered ages for the excluded calibration points are in very good agreement with the original calibrations, all of them falling within the calibration time frame set for that point, and close to the times obtained with all calibration points.

Applying these extended intervals, our results (Table 2 and Fig.4) indicated that the colonization events can be reliably dated into the Late Cretaceous–Middle Eocene for lemurs (70 to 41 Mya), Early Eocene–Early Miocene for tenrecs (50 to 20 Mya), and Early Oligocene–Middle Miocene for carnivorans and rodents (33 to 14 and 30 to 15 Mya, respectively). The time windows were synchronous for carnivorans and rodents, but there was no overlap between any of these two clades and the lemurs. The timing of the tenrec colonization overlapped in the Eocene with the lemurs and in the Oligocene-Miocene with the rodents and carnivorans. Because *Geogale*, possibly the most basal tenrec (Olson & Goodman 2003), was absent from our data set, the Malagasy tenrec radiation may actually be somewhat older and consequently their colonization window a bit narrower. In conclusion, Madagascar was colonized at a later period by carnivorans and rodents than by lemurs. Colonization by tenrecs may have occurred in the Late Eocene–Oligocene in a third, separate event, but we cannot exclude that it occurred simultaneously either with lemurs or with carnivorans and rodents.

Our dating of the lemur radiation at 50 Mya (59– 41 Mya) is more recent than the previous estimate by Yoder et al. (2003) at 66 Mya (75–55 Mya) using the same method, but agrees with a previous estimate of 48 to 41 Mya based on the epsilon-globin gene and its 5' flanking region (Porter et al. 1997; analyses performed with a local molecular clock approach). These differences could be due to the use of different phylogenetic markers (nuclear and mitochondrial genes) and to the fact that the IRBP gene (exon 1), used by Yoder et al. (2003), evolves significantly slower in lemurs than in other mammals, except perissodactyles (Poux et al. 2004). Our estimates for Malagasy carnivorans displayed a radiation time at 19 Mya (25–14 Mya), which is in perfect accordance with previous estimates (Yoder et al. 2003) of 20 Mya (26–15 Mya). Similarly, our datings for the split of nesomyine rodents and Malagasy tenrecs from their sister groups at 24 Mya (30–18 Mya) and 42 Mya (50–34 Mya), respectively, are not far frompreviously published results, 16±0.5/19 ± 1 Mya (Michaux et al. 2001; with global clock approach) and 43 Mya (52–34 Mya) (Douady & Douzery 2003; with Bayesian dating method), respectively.

To exclude the possibility that individual calibration constraints may bias our dating analyses, we repeated them after removing each calibration point in turn. All relevant datings remained highly congruent when any of the six calibrations was removed (Table 2). Moreover, the reciprocal compatibility of the calibrations was

evident: after excluding any of them, the remaining five calibrations always recovered a posterior estimate for the excluded node within the time window independently obtained from the corresponding fossil evidence (Table 2). In addition, the observed congruence of our interordinal divergence times with previously published data, based on much larger data sets, gives further confidence in our results (Table 3).

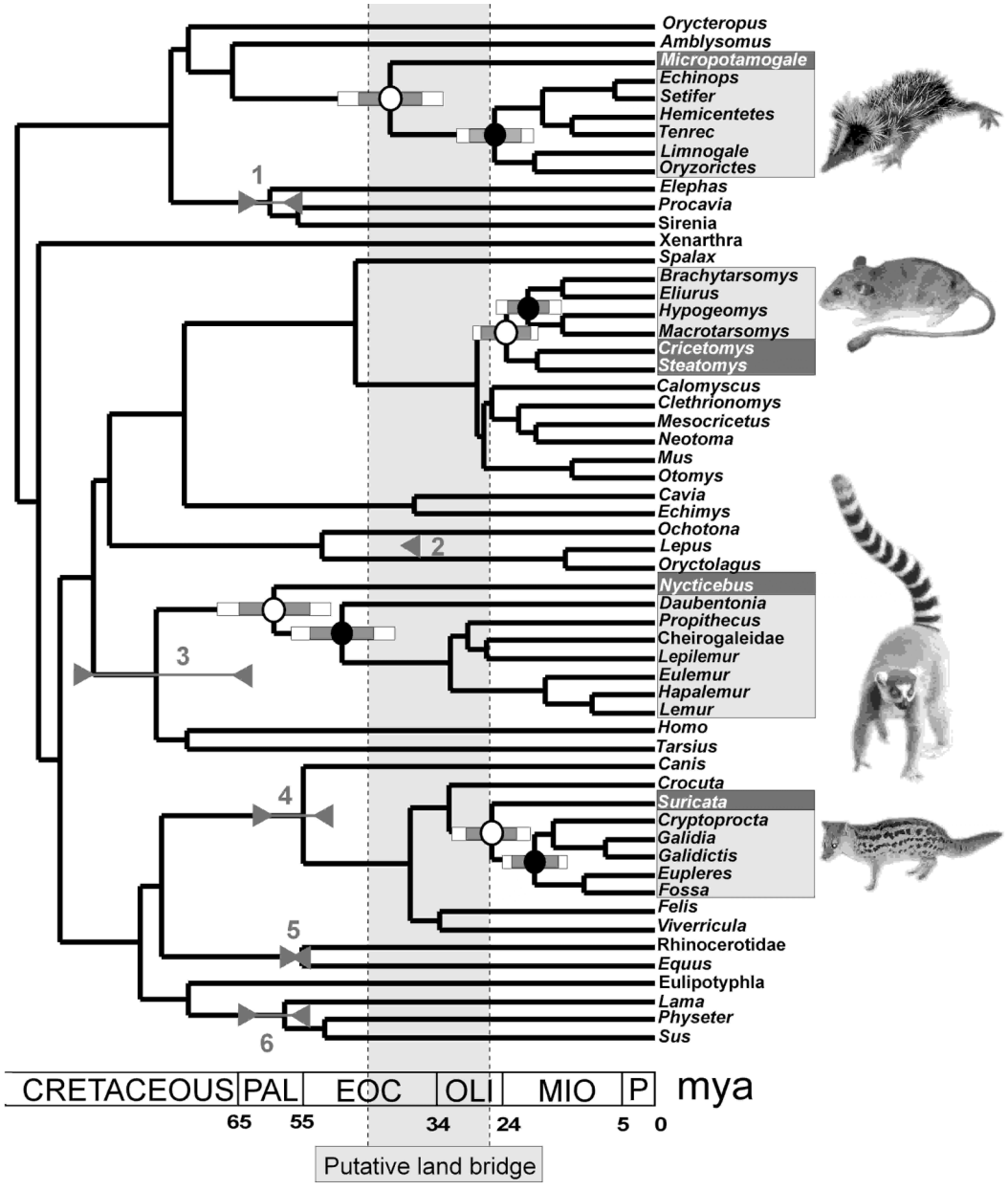

Figure 4: Asynchronous colonizations of the Malagasy mammal clades. Eutherian tree topology as in Figure 2. Divergence times were estimated from the concatenated data set by a Bayesian relaxed molecular clock method with six time constraints from fossil calibrations (nodes numbered as in Table 2). Malagasy clades are displayed in light grey boxes and their sister groups in dark grey boxes. Black circles indicate the initial divergence within each Malagasy radiation and open circles indicate divergences from non-Malagasy sister groups, with standard deviations (grey bars) and 95% credibility intervals (open bars) (see Table 2). Time estimates for all other nodes are given in Appendices 4 and 5 (available at www.systematicbiology.org). The period of a putative land bridge between Madagascar and Africa at 45 to 26 Mya (McCall, 1997) is shaded.

Table 3. Comparison of estimated divergence times (in Mya) with standard deviations (SD) and 95% credibility intervals (CI) from the present study, Springer et al. (2003), and Hasegawa et al. (2003). Divergence times in Springer et al. (2003) are based on a large data set (~16,000 bp) of mainly nuclear genes, and in Hasegawa et al. (2003) on a mitochondrial protein data set (3392 bp). Node numbers as in Appendices 4 and 5. In Hasegawa et al. (2003) the age of Laurasiatheria and 95% credibility intervals were not tabulated and could therefore not be included here (−). ND, not determined.

Clade and node number	This study		Springer et al. (2003)		Hasegawa et al. (2003)	
	age±SD	95% CI	age±SD	95% CI	age±SD	95% CI
Afrosoricida, 53	66.9±4.5	58.1-75.7	66.4±3.3	59.5-72.4	ND	ND
Afrotheria, 55	76.5±3.9	68.9-83.9	79.9±3.0	73.0-85.8	79.9±2.9	-
Glires, 41	86.3±4.6	77.3-95.4	82.6±3.2	76.6-89.0	74.6±1.6	-
Euarchontoglires, 42	89.0±4.4	80.4-97.8	87.3±3.2	81.5-93.9	89.0±1.9	-
Laurasiatheria, 15	81.6±3.3	75.3-88.4	85.1±2.5	80.3-90.3	-	-
Eutheria, 56 (root)	101.0±4.7	92.1-110.5	106.7±4.9	97.8-117.1	101.6±1.3	-

Biogeographic scenarios

In contrast to all other molecular studies of the mammalian colonization of Madagascar (Yoder et al. 1996, 2003; Michaux et al. 2001; Douady & Douzery 2003; Roos et al. 2004), we included all four Malagasy clades simultaneously in one analysis. Therefore, our estimates of divergence ages are directly comparable because they were affected by the same, if any, calibration biases. Our study is moreover based on the comparatively greatest length of concatenated nucleotides and includes representatives of 13 mammalian orders, which reduces sampling bias and long-branch attractions. The robustness of our results strengthens the evaluation of the different scenarios that have been proposed to explain the origin of extant Malagasy mammals: (i) ancient vicariance; (ii) terrestrial migration or island hopping along a land bridge or island arc; (iii) overseas rafting across the 400 km of open sea that make up the Mozambique channel.

The first scenario, vicariance, has been invoked for lemurs (Arnason et al. 2000) and would assume an age of colonization older than 84 Mya, the time when Madagascar became isolated (Briggs 2003). According to our data, lemurs were the first to diverge from their African sister group, not earlier than 70 Mya (including the 95% credibility interval). Vicariance can thus be excluded as an explanation for the origin of lemurs and any other Malagasy mammal lineage.

The second scenario involves a more or less continuous land bridge between Africa and Madagascar during the period 45 to 26Mya (McCall 1997). Our results do not match the colonization pattern expected under this hypothesis. Instead of showing

large overlapping periods between the four clades during the Middle Eocene–Late Oligocene, our results display colonization ages spread over the Tertiary (Fig.4). The radiation of lemurs dated at 75 to 55 Mya (Yoder et al. 2003) invalidated the land bridge hypothesis for this clade. However, in our study, the estimated age of the lemur colonization (70 to 41 Mya) is younger and therefore overlaps slightly with the postulated land bridge period. The windows of colonization (using the 95% credibility intervals) of tenrecs, carnivorans, and rodents likewise overlap to different extents the period of the putative land bridge, and migration via the land bridge route therefore cannot be excluded based on our data. However, the hypothesis remains unlikely because in three out of the four clades (all except tenrecs), both our ingroup and outgroup age estimates are outside of the landbridge period, the overlap only concerning the credibility intervals. Moreover, the existence of an emerged land bridge during the Eocene/Oligocene period has been seriously challenged (e.g. Rogers et al. 2000), and if this land bridge had been uninterrupted, a much greater variety of mammalian lineages could be expected to have colonized Madagascar.

The third scenario, transoceanic dispersal on rafting flotsam, predicts colonizations to occur probably randomly over time (Krause et al. 1997). The clearly asynchronous timing of at least two colonization events supports this scenario. Also considering that the estimated colonization times for lemurs, carnivorans, and rodents are largely outside the assumed time frame for the land bridge (with only the credibility intervals overlapping), we favor the transoceanic dispersal scenario. This agrees with the pattern observed in the majority of nonflying Malagasy vertebrate groups (Vences 2004) and in at least some plants (Yuan et al. 2005), and supports recent claims that the importance of oceanic dispersal has been strongly underestimated in historical biogeography (de Queiroz 2004).

In conclusion, the extant diversity of endemic Malagasy mammals reflects four adaptive radiations that probably colonized the island in at least two asynchronous waves of overseas dispersal. Studying ancient DNA from subfossil remains of two extinct lineages of Malagasy mammals, hippos and the enigmatic *Plesiorycteropus* (Goodman et al. 2003), bears the potential to add additional colonization ages and thereby test the hypothesis of random timing. Relating the age, pattern, and diversity of radiations to the emergence of eastern Malagasy rainforests in the Eocene or Oligocene (Wells 2003) is a further exciting perspective for studies on the Malagasy biota.

ACKNOWLEDGMENTS

We are grateful to numerous friends and colleagues who provided samples and/or useful information, in particular to F. Andreone, F. Catzeflis, E. J. P. Douzery, M. East, E. Edwards, H. Hofer, E. and T. Rajeriarison, J. Randrianirina, W. zum Vörde Sive Vörding, B. Wachter, C. Woodhead, and A. D. Yoder. Fieldwork was carried out in collaboration with the Département de Biologie Animale of the University of Antananarivo; we are indebted to O. Ramilijaona, D. Rakotondravony, and N. Raminosoa, and to the Malagasy authorities for permits. D.R.V. was supported by a grant of the University of Vigo, C.P. and O.M. were supported by grants from the Netherlands Organization of Scientific Research (NWO) and the European Commission.

REFERENCES

Arnason, U., A. Gullberg, A. S. Burguete, and A. Janke. (2000). Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* 133:217–228.

Asher, R. J. (1999). Amorphological basis for assessing the phylogeny of the "Tenrecoidea" (Mammalia, Lipotyphla). *Cladistics* 15:231–252.

Benton, M. J. (1993). Fossil record 2. Chapman and Hall, London.

Bossuyt, F., and M. C. Milinkovitch. (2001). Amphibians as indicators of early tertiary "out-of-India" dispersal of vertebrates. *Science* 292:93– 95.

Briggs, J. C. (2003). The biogeographic and tectonic history of India. *J. Biogeogr.* 30:381–388.

Bromham, L., and D. Penny. (2003). The modern molecular clock. *Nat. Rev. Genet.* 4:216–224.

De Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends Ecol. Evol.* 20:68–73.

Douady, C. J., and E. J. Douzery. (2003). Molecular estimation of eulipotyphlan divergence times and the evolution of "Insectivora." *Mol. Phylogenet. Evol.* 28:285–296.

Dubois, J.-Y., D. Rakotondravony, C. Hänni, P. Sourrouille, and F. Catzeflis. (1996). Molecular evolutionary relationships of three genera of Nesomyinae, endemic rodent taxa from Madagascar. *J. Mamm. Evol.* 3:239–260.

Eisenberg, J. F. (1981). The mammalian radiations. An analysis of trends in evolution, adaptation and behavior. The University of Chicago Press, Chicago and London.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.

Garland, T. J., A. W. Dickerman, C. M. Janis, and J. A. Jones. (1993). Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42:265–292.

Gatesy, J., and M. A. O'Leary. (2001). Deciphering whale origins with molecules and fossils. *Trend. Ecol. Evol.* 16:562–570.

Gheerbrant, E., J. Sudre, M. Larochene, and A. Moumni. (2001). First ascertained African "Condylarth" mammals (Primitive ungulates: cf. Bulbulodentata and cf. Phenacodonta) from the earliest Ypresian of the Ouled Abdoun Basin, Morocco. *J. Vert. Paleont.* 21:107–118.

Gingerich, P. D., and M. D. Uhen. (1994). Time of origin of primates. *J. Hum. Evol.* 27:443–445.

Glazko, G. V., and M. Nei. (2003). Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20:424–434.

Goodman, S. M., J. U. Ganzhorn, and D. Rakotondravony. (2003). Introduction to the Mammals. Pages 1159–1186 *in* The natural history of Madagascar (S. M. Goodman, and J. P. Benstead, eds.). Chicago University Press, Chicago.

Grant, P. R. (1998). Evolution on islands. Oxford University Press, Oxford.

Graur, D., and W. Martin. (2004). Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20:80–86.

Hasegawa, M., J. L. Thorne, and H. Kishino. (2003). Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78:267–283.

Hedges, S. B., and S. Kumar. (2003). Genomic clocks and evolutionary timescales. *Trends Genet.* 19:200–206.

Ho, S. Y. W., M. J. Phillips, A. J. Drummond, and A. Cooper. (2005). Accuracy of rate estimation using relaxed clock models, with a critical focus on the early metazoan radiation. *Mol. Biol. Evol.* 22:1355–1363.

Huelsenbeck, J. P., and F. Ronquist. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Jansa, S. A., S. M. Goodman, and P. K. Tucker. (1999). Molecular phylogeny and biogeography of the native rodents of Madagascar (Muridae: Nesomyinae): A test of the single-origin hypothesis. *Cladistics* 15:253–270.

Jansa, S. A., and M.Weksler. (2004). Phylogeny of muroid rodents: Relationships within and among major lineages as determined by IRBP gene sequences. *Mol. Phylogenet. Evol.* 31:256–276.

Krause, D.W. (2001). Fossil molar fromaMadagascan marsupial. *Nature* 412:497–498.

Krause, D. W., J. H. Hartman, and N. A. Wells. (1997a). Late Cretaceous vertebrates from Madagascar. Implications for biotic changes in deep time. Pages 3–43 in Natural change and human impact in Madagascar (B. D.GoodmanandS.M.Patterson, eds.) Smithsonian Institution Press, Washington, DC.

Krause, D.W., G. V. R. Prasad,W. von Koenigswald, A. Sahni, and F. E. Grine. (1997b). Cosmopolitanism among Gondwanan Late Cretaceous mammals. *Nature* 390:504–507.

Lavocat, R. (1978). Rodentia and Lagomorpha. Pages 69–89 *in* Evolution of African mammals (V. J.Maglio and H. B. S. Cooke, eds.). Harvard University Press, Boston.

Lawlor, T. E. (1986). Comparative biogeography of mammals on islands. *Biol. J. Linnean Soc.* 28:99–125.

Losos, J. B., T. R. Jackman, A. Larson, K. Queiroz, and L. Rodriguez-Schettino. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279:2115–2118.

Martin, R. D. (1993). Primate origins: Plugging the gaps. *Nature* 363:223– 234.

McCall, R. A. (1997). Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc. R. Soc. Lond. B Biol. Sci.* 264:663–665.

McKenna, M. C., and S. K. Bell. (1997). Classification of mammals above the species level. Columbia University Press, New York.

Michaux, J., A. Reyes, and F. Catzeflis. (2001). Evolutionary history of the most speciose mammals: Molecular phylogeny of muroid rodents. *Mol. Biol. Evol.* 18:2017–2031.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, and J. Kent. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.

Nagy, Z. T., U. Joger, M. Wink, F. Glaw, and M. Vences. (2003). Multiple colonization of Madagascar and Socotra by colubrid snakes: Evidence from nuclear and mitochondrial gene phylogenies. *Proc. R. Soc. Lond. B Biol. Sci.* 270:2613–2621.

Olson, L. E., and S. M. Goodman. (2003). Phylogeny and biogeography of Tenrecs. Pages 1235–1242 *in* The natural history of Madagascar (S. M. Goodman, and J. P. Benstead, eds.). Chicago University Press, Chicago.

Pastorini, J., U. Thalmann, and R. D. Martin. (2003). A molecular approach to comparative phylogeography of extant Malagasy lemurs. *Proc. Natl. Acad. Sci. USA.* 100:5879–5884.

Porter,A. P., M. Goodman, and M. J. Stanhope. (1996). Evidence on mammalian phylogeny from sequences of exon 28 of the vonWillebrand Factor gene. *Mol. Phylogenet. Evol.* 5:89–101.

Porter, C. A., S. L. Page, J. Czelusniak, I. Sampaio, H. Schneider, M. P. C. Schneider, I. Sampaio, and M. Goodman. (1997). Phylogeny and evolution of selected Primates as determined by sequences of the ε-globin locus and 5_ flanking regions. *Int. J. Prim.* 18:261–295.

Posada, D., and K. A. Crandall. (1998). ModelTest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.

Poux, C., P. Chevret, D. Huchon, W. W. de Jong, and E. J. P. Douzery. In press. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.*

Poux, C., and E. J. P. Douzery. (2004). Primate phylogeny, evolutionary rate variation, and divergence times:Acontribution fromthe nuclear gene IRBP. *Am. J. Phys. Antropol.* 124:1–16.

Rogers, R. R., J. H. Hartman, and D. W. Krause. (2000). Stratigraphic analysis of Upper Cretaceous rocks in the Mahajanga Basin, northwestern Madagascar: Implications for ancient and modern faunas. *J. Geol.* 108:275–301.

Roos, C., J. Schmitz, and H. Zischler. (2004). Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl. Acad. Sci. USA* 101:10650– 10654.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.* 19:101–109.

Seiffert, E. R., E. L. Simons, and Y. Attia. (2003). Fossil evidence for an ancient divergence of lorises and galagos. *Nature* 422:421–424.

Springer,M. S., G. C. Cleven, O. Madsen,W.W. de Jong, V. G.Waddell, H. M. Amrine, and M. J. Stanhope. (1997). Endemic african mammals shake the phylogenetic tree. *Nature* 388:61–64.

Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope. (2001). Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18:132–143.

Springer,M. S.,W. J.Murphy,E. Eizirik, and S. J. O'Brien. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100:1056–1061.

Springer, M. S., M. J. Stanhope, O. Madsen, and W. W. de Jong. (2004). Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* 19:430–438.

Steppan, S. J., R. M. Adkins, and J. Anderson. (2004). Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst. Biol.* 53:533–553.

Swofford, D. L. (2003). PAUP*. Phylogenetic analysis using parsimony (∗ and other methods). version 4. Sinauer Associates, Sunderland, Massachusetts.

Tavare, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. (2002). Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* 416:726–729.

Thorne, J. L., and H. Kishino. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–1657.

Vences, M. (2004). Origin of Madagascar's extant fauna: A perspective from amphibians, reptiles and other non-flying vertebrates. *Ital. J. Zool.* (Suppl.) 2:217–228.

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer. (2003). Multiple overseas dispersal in amphibians. *Proc. R. Soc. Lond. B Biol. Sci.* 270:2435–2442.

Wells, N. A. (2003). Some Hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar. Pages 16–34 *in* The natural history of Madagascar (S. M. Goodman, and J. P. Benstead, eds.). Chicago University Press, Chicago.

Yoder, A. D., M. M. Burns, S. Zehr, T. Delefosse, G. Veron, S. M. Goodman, and J. J. Flynn. (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421:734–737.

Yoder, A. D., M. Cartmill, M. Ruvolo, K. Smith, and R. Vilgalys. (1996). Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. USA* 93:5122–5126.

Yuan, Y.-M., S. Wohlhauser, M. Möller, J. Klackenberg, M. W. Callmander, and P.Küpfer. (2005). Phylogeny and biogeography of *Exacum* (Gentianaceae): A disjunctive distribution in the Indian Ocean basin resulted from long distance dispersal and extensive radiation. *Syst. Biol.* 54:21–34.

Appendix 1. Taxonomic sampling and accession numbers for the three nuclear genes. Upperscore numbers (1−10) refer to data sets in which different taxa were available for each gene, and were concatenated, or to taxa that were not included in all analyses. ∗Sequences taken from the database.

	Species	ADRA2B	vWF	AR
Eutheria				
RODENTIA				
Sciurognathi				
Muridae				
Murinae	*Mus musculus*	L00979*	AJ238390*	NM_013476*
Spalacinae	*Spalax ehrenbergii*	AJ891078	U31621*	AJ893519
Nesomyinae	*Eliurus* sp.	AJ891058	AJ891086	AJ893520
	Hypogeomys antimena	AJ891066	AJ891094	AJ893521
	Macrotarsomys ingens	AJ891070	AJ402705*	AJ893522
	Brachytarsomys albicauda	AJ891049	AJ891083	AJ893523
Arvicolinae	*Clethrionomys glareolus*	AJ891053	AJ402709*	AJ893524
Calomyscinae	*Calomyscus mystax*	AJ891050	AJ402702*	AJ893525
Cricetinae	*Mesocricetus auratus*	AJ891071	AJ402706*	AJ893526
Cricetomyinae	*Cricetomys gambianus*	AJ891054	AJ402694*	AJ893527
Dendromurinae	*Steatomys* cf. *gautuni*	AJ891079	AJ402704*	AJ893528
Otomyinae	*Otomys angoniensis*	AJ891075	AJ402711*	AJ893529
Sigmodontinae	*Neotoma fuscipes*	AJ891073	AJ402703*	AJ893530
Hystricognathi				
Echimyidae	*Echimys chrysurus*	AJ427269*	AJ251141*	AJ893532
Caviidae	*Cavia porcellus*	AJ271336*	AJ224663*	AJ893531
LAGOMORPHA				
Leporidae	*Oryctolagus cuniculus*	Y15946*	U31618*	AJ893533
	Lepus crawshayi	AJ427254*	AJ224669*	AJ893534
Ochotonidae	*Ochotona princeps*	AJ427253*	AJ224672*	AJ893535
PRIMATES				
Lemuridae	*Lemur catta*	AJ891067	AJ410292*	AJ893536
	Eulemur fulvus fulvus	AJ891059	AJ891087	AJ893537
	Hapalemur simus	AJ891064	AJ891092	AJ893538
Megaladapidae	*Lepilemur edwardsi*	AJ891068	AJ891095	AJ893539
Cheirogaleidae	*Cheirogaleus / Microcebus*[1]	AJ891052	AJ410295*	AJ893540
Daubentoniidae	*Daubentonia madagascariensis*	AJ891057	AJ410293*	AJ893541
Indridae	*Propithecus verreauxi coronatus*	AJ891076	AJ410294*	AJ893542
Loridae	*Nycticebus coucang*	AJ251186*	AJ410291*	AJ893543
Tarsiidae	*Tarsius bancanus*	AJ891081	AJ410296*	AJ893544
Hominidae	*Homo sapiens*	M34041*	X06828 *	M27423*
CARNIVORA				
Canidae	*Canis familiaris*	AJ891051	L16903*	AF197950*
Felidae	*Felis catus*	AJ251174*	U31613*	AJ893545
Hyaenidae	*Crocuta crocuta*	AJ891055	AJ891084	AY128705*
Herpestidae / Viverridae				
Galidiinae	*Galidictis fasciata*	AJ891063	AJ891091	AJ893547
	Galidia elegans	AJ891062	AJ891090	AJ893546
Herpestinae	*Suricata suricata*	AJ891080	AJ891099	AJ893548
Cryptoproctinae	*Cryptoprocta ferox*	AJ891056	AJ891085	AJ893549
Euplerinae	*Eupleres goudoti*	AJ891060	AJ891088	AJ893550
	Fossa fossana	AJ891061	AJ891089	AJ893551
Viverrinae	*Viverricula indica*	AJ891082	AJ891100	AJ893552
PERISSODACTYLA				
Rhinocerotidae	*Ceratotherium / Diceros*[2]	AJ251184*	U31604*	AJ893553
Equidae	*Equus* sp.[3]	Y15945*	U31610*	AJ893554
CETARTIODACTYLA				
Camelidae	*Lama*[4]	AJ315941*	AF108835 *	AJ893555
Suidea	*Sus scrofa*	AJ251177*	S78431*	AF161717*

	Species	ADRA2B	vWF	AR
Physeteridae	*Physeter catodon*	AJ427417*	AF108834*	AJ893556
EULIPOTYPHLA	*Erinaceus / Crocidura*[5]	Y12521*	AY057834*	AJ893557
XENARTHRA	*Bradypus / Cyclopes*[6]	AJ251179*	U31603*	AJ893558
SIRENIA	*Trichechus / Dugong*[7]	AJ251109*	U31608*	AJ893559
PROBOSCIDEA	*Elephas maximus*	Y12525*	U31611*	AJ893560
HYRACOIDEA	*Procavia capensis*	Y12523*	U31619*	AJ893561
TUBULIDENTATA	*Orycteropus afer*	Y12522*	U31617*	AJ893563
AFROSORICIDA				
Chrysochloridea	*Amblysomus hottentotus*	Y12526*	U97534*	AJ893562
Tenrecidea				
Tenrecinae	*Setifer setosus*	AJ891077	AJ891098	AJ893566
	Echinops telfairi	Y17692*	AF076478*	AJ893565
	Tenrec ecaudatus	AJ251108*	AF390536*	AJ893564
	Hemicentetes semispinosus	AJ891065	AJ891093	AJ893567
Oryzoryctinae	*Oryzorictes hova*	AJ891074	AJ891097	AJ893568
	Microgale brevicaudata[8]	AJ891072	-	AJ893569
	Limnogale mergulus	AJ891069	AJ891096	AJ893570
Potamogalinae	*Micropotamogale lamottei*	AJ251107*	AF390538*	AJ893571
Marsupialia				
DIDELPHIMORPHIA	*Didelphis*[9]	Y15943*	AF226848*	AJ893572
DIPROTODONTIA	*Macropus*[10]	AJ251183*	AJ224670*	AJ893573

[1]*Cheirogaleus medius* (ADRA2B, AR) combined with *Microcebus murinus* (vWF).
[2]*Diceros bicornis* (ADRA2B, AR) combined with *Ceratotherium simum* (vWF).
[3]*Equus caballus* (ADRA2B, AR) combined with *E. asinus* (vWF).
[4]*Lama pacos* (ADRA2B, AR) combined with *L. glama* (vWF).
[5]*Erinaceus europaeus* (ADRA2B, AR) combined with *Crocidura russula* (vWF).
[6]*Bradypus tridactylus* (ADRA2B, vWF) combined with *Cyclopes didactylus* (AR).
[7]*Trichechus manatus* (ADRA2B, AR) combined with *Dugong dugon* (vWF).
[8]*Microgale brevicaudata* was removed from the dating analyses.
[9]*Didelphis marsupialis* (ADRA2B, AR) combined with *D. virginiana* (vWF).
[10]*Macropus rufus* (ADRA2B, AR) combined with *M. giganteus* (vWF).

Chapter 6

Molecular phylogeny and divergence times of Malagasy tenrecs: Influence of data partitioning and taxon sampling on dating analyses

Céline Poux[1,3], Ole Madsen[1,4], Julian Glos[2,5], Wilfried W. de Jong[1] and Miguel Vences[2]

[1] Department of Biomolecular Chemistry 271, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

[2] Division of Evolutionary Biology, Zoological Institute, Technical University of Braunschweig, Braunschweig, Germany.

[3] Vertebrate Department, Royal Belgian Institute of Natural Sciences, Brussels, Belgium.

[4] Animal Breeding and Genomics Center, Wageningen University, Wageningen, The Netherlands.

[5] Animal Ecology and Conservation Biology Department, Biocenter Grindel and Zoological Museum, Hamburg, Germany.

BMC Evolutionary Biology 8:102, 2008

ABSTRACT

Malagasy tenrecs belong to the Afrotherian clade of placental mammals and comprise three subfamilies divided in eight genera (Tenrecinae: *Tenrec*, *Echinops*, *Setifer* and *Hemicentetes*; Oryzorictinae: *Oryzorictes*, *Limnogale* and *Microgale*; Geogalinae: *Geogale*). The diversity of their morphology and incomplete taxon sampling made it difficult until now to resolve phylogenies based on either morphology or molecular data for this group. Therefore, in order to delineate the evolutionary history of this family, phylogenetic and dating analyses were performed on a four nuclear genes dataset (ADRA2B, AR, GHR and vWF) including all Malagasy tenrec genera. Moreover, the influence of both taxon sampling and data partitioning on the accuracy of the estimated ages were assessed.

Within Afrotheria the vast majority of the nodes received a high support, including the grouping of hyrax with sea cow and the monophyly of both Afroinsectivora (Macroscelidea + Afrosoricida) and Afroinsectiphillia (Tubulidentata + Afroinsectivora). Strongly supported relationships were also recovered among all tenrec genera, allowing us to firmly establish the grouping of *Geogale* with Oryzorictinae, and to confirm the previously hypothesized nesting of *Limnogale* within the genus *Microgale*. The timeline of Malagasy tenrec diversification does not reflect a fast adaptive radiation after the arrival on Madagascar, indicating that morphological specializations have appeared over the whole evolutionary history of the family, and not just in a short period after colonization. In our analysis, age estimates at the root of a clade became older with increased taxon sampling of that clade. Moreover an augmentation of data partitions resulted in older age estimates as well, whereas standard deviations increased when more extreme partition schemes were used.

Our results provide as yet the best resolved gene tree comprising all Malagasy tenrec genera, and may lead to a revision of tenrec taxonomy. A timeframe of tenrec evolution built on the basis of this solid phylogenetic framework showed that morphological specializations of the tenrecs might have been affected by environmental changes caused by climatic and/or subsequent colonization events. Analyses including various taxon sampling and data partitions allow us to point out

some possible pitfalls that may lead to biased results in molecular dating; however, further analyses are needed to corroborate these observations.

INTRODUCTION

The Malagasy tenrecs belong to the Afrotheria, one of the four basal clades of placental mammals which have recently been recognized (Murphy et al. 2001). This ancient group of African origin is divided into two clades: the strongly supported Paenungulata, composed of the orders Sirenia (sea cows), Proboscidea (elephants) and Hyracoidea (hyraxes), and the Afroinsectiphillia (Waddell et al. 2001), comprising the orders Afrosoricida (golden moles and tenrecs), Macroscelidea (elephant shrews) and Tubulidentata (aardvark) (Springer et al. 2004; Nishihara et al. 2005). The tenrec family (Tenrecidae) comprises four subfamilies, the Potamogalinae from continental Africa, and the Tenrecinae, Geogalinae and Oryzorictinae from Madagascar. The Malagasy tenrecs are divided into eight genera and 30 species (Bronner & Jenkins 2005; Olson et al. 2004; Goodman et al. 2006; Goodman & Jenkins 2007). Based on morphology, tenrecs were previously grouped in the insectivorous order Lipotyphla, which has turned out to be biphyletic and now is split into the orders Eulipotyphla (hedgehogs, moles, shrews, solenodons) and Afrosoricida (Stanhope et al. 1998)

The Malagasy tenrecs have diversified into a spectacular radiation in terms of morphology, behavior, physiology and ecology. They show a high degree of adaptation to their niches (terrestrial, semi-arboreal, fossorial and semiaquatic) and considerable convergence with other insectivores, notably shrews and hedgehogs. This made it difficult to understand the origin and phylogenetic relationships of this group on a morphological basis. The Tenrecinae (spiny tenrecs) include four genera (*Hemicentetes*, *Tenrec*, *Setifer*, *Echinops*), characterized by a spiny pelage and a large body size compared to the other tenrecs. Their monophyly is well established, even at the morphological level (Asher 1999). The branching of the four remaining genera (*Geogale*, *Oryzorictes*, *Limnogale* and *Microgale*), which share a shrew-like appearance and a small size, remains more open. Most earlier, molecular studies did not include more than five tenrec species (Emerson et al. 1999; Mouchaty et al. 2000, Douady et al. 2002, Douady & Douzery 2003, Malia et al. 2002), while Poux et al. (2005) missed the large-eared tenrec (*Geogale*). Therefore, not all relations between

and within the three subfamilies of Malagasy tenrecs have yet been firmly established. Only two recent studies, by Olson & Goodman (2003) and Asher & Hofreiter (2006), included all tenrec genera, but were unable to confidently resolve the position of *Geogale*, which suggests the necessity to expand the number of species and sequences for this family.

The island of Madagascar is a well-known biodiversity hotspot, displaying diverse and highly endemic amphibian, reptilian and mammalian faunas. The level of endemism reaches 95% for the non-flying vertebrates, and this level is mainly due to a few speciose endemic radiations (Bossuyt & Milinkovitch 2001; Nagy et al. 2003; Vences et al. 2003). Four clades of terrestrial endemic mammals are present, the lemuriform primates, the euplerine carnivores, the nesomyine rodents and the Malagasy tenrecs. Each of these clades represents one unique event of colonization from continental Africa, followed by several diversification events that gave rise to the actual Malagasy diversity (Poux et al. 2005; Yoder et al. 2003). The colonization of a new environment can be followed by an adaptive radiation, defined as a rapid succession of speciation events leading to a high ecological and phenotypic diversity within a lineage (Schluter 2000). The study of adaptive radiations on islands or in lakes is essential for understanding processes of speciation and diversification (Grant 1998; Losos et al. 1998; Seehausen 2006). Therefore, knowing the patterns and timing of the successive diversification events within endemic island clades, which, like tenrecs, display a broad ecological and morphological diversity, might help to better understand this phenomenon.

Apart from *Echinops telfairi*, for which the genome sequencing is in progress, there are only a limited number of sequences available in public databases to reconstruct a solid molecular phylogeny of the Malagasy tenrecs. In the present study we therefore selected exons from four independent nuclear genes that are widely used in mammalian phylogeny (ADRA2B, AR, GHR and vWF) in order to resolve tenrec phylogeny. This study is especially focused on understanding the phylogenetic position of the largeeared and the web-footed tenrecs, *Geogale* and *Limnogale*, respectively. In addition, we used a relaxed molecular clock timeframe to compare tenrec evolutionary patterns with defined adaptive radiation characteristics. Moreover, the influence of both taxon sampling and data partitioning on the accuracy of the estimated ages were assessed.

MATERIAL AND METHODS

Sampling, DNA amplification and sequencing

Fragments of the intronless gene of the alpha 2B adrenergic receptor (ADRA2B), of exon 1 of the androgen receptor (AR) gene, of exon 10 of the growth hormone receptor (GHR) gene, and of exon 28 of the von Willebrand factor (vWF) gene were amplified and sequenced. These genes were selected because (i) they are located in the nuclear genome, as single-copy genes (in at least human and mouse), (ii) a considerable number of sequences are already available for all four genes and have been useful in mammalian phylogeny, and (iii) they are functionally and genetically unrelated. We selected for each of the four genes 38 mammalian species to represent (i) all genera of Malagasy tenrecs, and at least two species of the very diverse genus *Microgale*, in order to assess the phylogenetic position of *Limnogale*, (ii) the continental African sister group (Potamogalinae) of the Malagasy tenrecs, (iii) groups needed for multiple calibrations of the molecular clock, (iv) at least one species from each eutherian order (but for Pholidota), and (v) appropriate marsupial outgroups. A total of 19 new sequences were obtained, and complemented with 134 sequences from GenBank (Table 1).

Genomic DNA was isolated from ethanol-preserved tissue, following the protocols of the Wizard® SV Genomic DNA Purification System (Promega). Fragments of the ADRA2B and AR genes were amplified using previously published primers (Poux et al. 2005; Springer et al. 1997). New primers were designed for vWF and GHR. For these last genes PCR reactions were performed on 50–200 ng DNA with Expand DNA polymerase (Expand High Fidelity PCR system, Roche) using the following program: 2 min at 94°C; 30–35 cycles of 15 sec at 94°C, 1 min at 60°C and 1 min 30 sec at 72°C; and a final step of 2–10 min at 72°C. DMSO (1.3 – 2.5%) and/or betaine (1 M) was added for some samples. PCR products were purified from a 1% agarose gel, using GFX™ PCR DNA & Gel Band Purification Kit (GE Healthcare), and reamplified if necessary. Gel-extracted PCR products were sequenced directly on a 3730 96-capillary sequencer (Applied Biosystems). Internal primers were used to get complete sequences of both strands.

Table 1: Taxonomic sampling and accession numbers of the four nuclear genes. Upperscore numbers ([1–14]) refer to taxa for which sequences from different species were combined in the concatenated analysis. * New sequences from the present study. The full alignment is available from Treebase (accession number M3679).

	Species	ADRA2B	AR	GHR	vWF
Eutheria					
RODENTIA					
Muridae	*Mus musculus*	L00979	NM_013476	M33324	AJ238390
Caviidae	*Cavia porcellus*	AJ271336	AJ893531	AF238492	AJ224663
Sciuridae	*Marmota /Sciurus* [1]	AJ315942	AM905334*	AF332032	J224671
LAGOMORPHA					
Leporidae	*Oryctolagus cuniculus*	Y15946	AJ893533	AF015252	U31618
	Lepus sp. [2]	AJ427254	AJ893534	AF332016	AJ224669
Ochotonidae	*Ochotona princeps*	AJ427253	AJ893535	AF332015	AJ224672
PRIMATES					
Lemuridae	*Eulemur sp.* [3]	AJ891059	AJ893537	AF540627	AJ891087
Hominidae	*Homo sapiens*	M34041	M27423	X06562	X06828
SCANDENTIA	*Tupaia sp.* [4]	AJ251187	AM905335*	AF540643	U31624
DERMOPTERA	*Cynocephalus variegatus*	AJ251182	AM905340*	AF540625	U31606
CARNIVORA					
Canidae	*Canis familiaris*	AJ891051	AF197950	AF133835	L16903
Felidae	*Cryptoprocta ferox*	AJ891056	AJ893549	AY928733	AJ891085
PERISSODACTYLA					
Rhinocerotidae	*Ceratotherium / Diceros* [5]	AJ251184	AJ893553	AM905343*	U31604
Equidae	*Equus sp.* [6]	Y15945	AJ893554	AF392878	U31610
CETARTIODACTYLA					
Camelidae	*Lama sp.* [7]	AJ315941	AJ893555	AM905349*	AF108835
Suidae	*Sus scrofa*	AJ251177	AF161717	X54429	S78431
Physeteridae	*Physeter catodon*	AJ427417	AJ893556	AM905344*	AF108834
CHIROPTERA	*Cynopterus / Pteropus* [8]	AJ251181	AM905339*	AF392893	U31605
EULIPOTYPHLA	*Erinaceus / Crocidura* [9]	Y12521	AJ893557	AF392882	AY057834
XENARTHRA	*Myrmecophaga / Cyclopes* [10]	MTR427373	AJ893558	AF392875	MTR278157
SIRENIA	*Trichechus / Dugong* [11]	AJ251109	AJ893559	AF392891	U31608
PROBOSCIDEA	*Elephas maximus*	Y12525	AJ893560	AF332013	U31611
HYRACOIDEA	*Procavia capensis*	Y12523	AJ893561	AF392896	U31619
TUBULIDENTATA	*Orycteropus afer*	Y12522	AJ893563	AF392892	U31617
MACROSCELIDEA	*Macroscelides proboscideus*	Y12524	AM905337*	AF332014	AY310893
AFROSORICIDA					
Chrysochloridae	*Amblysomus / Chrysospalax* [12]	Y12526	AJ893562	AF392877	U97534
Tenrecidae					
Tenrecinae	*Setifer setosus*	AJ891077	AJ893566	DQ202292	AJ891098
	Echinops telfairi	Y17692	AJ893565	AF392889	AF076478
	Tenrec ecaudatus	AJ251108	AJ893564	AF392890	AF390536
	Hemicentetes semispinosus	AJ891065	AJ893567	DQ202288	AJ891093
Oryzoryctinae	*Oryzorictes hova*	AJ891074	AJ893568	AF392886	AJ891097
	Microgale talazaci	-	-	AF392885	-
	Microgale brevicaudata	AJ891072	AJ893569	AM905345*	AM905350*
	Microgale cf. parvula	AM905341*	AM905336*	AM905346*	AM905351*
	Limnogale mergulus	AJ891069	AJ893570	DQ202289	AJ891096
Geogalinae	*Geogale aurita*	AM905342*	AM905338*	AM905347*	AM905352*
Potamogalinae	*Micropotamogale lamottei*	AJ251107	AJ893571	DQ202290	AF390538
Marsupialia					
DIDELPHIMORPHIA	*Didelphis / Monodelphis* [13]	Y15943	AJ893572	AF238491	AF226848
DIPROTODONTIA	*Macropus sp.* [14]	AJ251183	AJ893573	AM905348*	AJ224670

[1] *Sciurus vulgaris* (ADRA2B, AR) combined with *S. niger* (GHR) and *Marmota monax* (vWF)
[2] *Lepus crawshayi* (ADRA2B, AR, vWF) combined with *L. capensis* (GHR)
[3] *Eulemur fulvus fulvus* (ADRA2B, AR, vWF) combined with *E. coronatus* (GHR)
[4] *Tupaia tana* (ADRA2B, AR, GHR) combined with *T. glis* (VWF)

[5] *Diceros bicornis* (ADRA2B, AR, GHR) combined with *Ceratotherium simum* (vWF)
[6] *Equus caballus* (ADRA2B, AR, GHR) combined with *E. asinus* (vWF)
[7] *Lama pacos* (ADRA2B, AR, GHR) combined with *L. glama* (vWF)
[8] *Cynopterus sphinx* (ADRA2B, AR, vWF) combined with *Pteropus vampyrus* (GHR)
[9] *Erinaceus europaeus* (ADRA2B, AR, GHR) combined with *Crocidura russula* (vWF)
[10] *Myrmecophaga tridactyla* (ADRA2B, vWF, GHR) combined with *Cyclopes didactylus* (AR)
[11] *Trichechus manatus* (ADRA2B, AR, GHR) combined with *Dugong dugon* (vWF)
[12] *Amblysomus hottentotus* (ADRA2B, AR, vWF) combined with *Chrysospalax trevelyani* (GHR)
[13] *Didelphis marsupialis* (ADRA2B, AR) combined with *D. virginiana* (vWF) and *Monodelphis domestica* (GHR)
[14] *Macropus rufus* (ADRA2B, AR, GHR) combined with *M. giganteus* (vWF)

Phylogenetic analyses

Sequences were assembled and aligned with the ED editor of the MUST package (Philippe 1993), and manually adjusted taking amino acid properties in consideration. Amino acid repeats and sites not sequenced or gapped in more than 25% of the taxa were excluded from analysis. This resulted in a dataset of 1,101 bp for ADRA2B, 1,161 bp for AR, 852 bp for GHR, and 1,173 bp for vWF. The full data matrix is available from Treebase (accession number: M3679). Phylogenetic reconstructions on each gene separately and on the concatenated dataset were performed by maximum likelihood (ML) with PAUP*, version 4b10 (Swofford 2003), and by Bayesian analyses with MRBAYES, version 3.1.2 (Huelsenbeck & Ronquist 2001). The best fitting model under the ML criterion was selected from the "Akaike Criterion" output of MODELTEST, version 3.7 (Posada & Crandall 1998). The ML analysis was conducted using a loop approach to estimate the best tree and the optimal likely hood parameters. With this approach parameters and best tree are re-estimated until they reach stability. Node stability was estimated by 100 non-parametric bootstrap replicates (Felsenstein 1985). A major advantage of Bayesian phylogenetic inference is the possibility of partitioning the data, giving each partition its own best fitting model of sequence evolution. However, overpartitioning may introduce unnecessary sampling variances which could influence the phylogenetic estimates. For the twelve possible codon partitions (each codon position of each gene) MODELTEST was used to calculate the best fitting model of sequence evolution. As further explained in Table 2, codon partitions with similar models and model parameters were merged, resulting in nine partitions for the Bayesian analyses. Two runs of four Markov chains were calculated simultaneously for 1,000,000 generations with initial equal probabilities for all trees and starting with a random tree. Tree

sampling frequency was each 20 generations, and the consensus tree with posterior probabilities was calculated after removal of the first 25% of the total number of trees generated, corresponding to 12,500 trees. The average standard deviation of split frequencies between the two independent runs was lower than 0.01.

Table 2: Best fitting evolutionary model for each codon position. Best models and parameters were found with the akaike criterion as implemented in MODELTEST 3.7 and with PAML, for each codon position of the four gene fragments. Codon positions with similar model and model parameters were regrouped into the same partition, which resulted in nine partitions when estimated by MODELTEST and five partitions when estimated by PAML. Codon positions were merged into the same partition when none of their model parameters (e.g., TRatio of position 1 compared to TRatio of position 2, PInvar 1 to PInvar 2, etc.) differed by more than 100%. For the parameters estimated by PAML we took also into account, to define the partitions, the rate of the various gamma low categories; these parameters are not included in this table. TRatio, transition/transversion ratio; Rmat, rate matrix; π, base frequency; PInvar, proportion of invariable sites; alpha, shape of gamma distribution; kappa, value of the transition/transversion ratio under the F84 model. CP stands for codon position and PN for partition number

						Estimated by MODELTEST					Estimated by PAML		
Gene	Codon position	Length	π_A	π_C	π_G	Best model	TRatio or Rmat	alpha	PInvar	Partition Number	kappa	alpha	Partition Number
ADRA2B	1	367	0.22	0.31	0.28	K81uf+I+Γ	(1.0 2.5 0.7 0.7 2.5)	1.04	0.36	1	1.05	0.34	1
	2	367	0.19	0.30	0.21	GTR+Γ	(1.6 6.1 0.7 2.6 3.6)	0.24	0	2	1.18	0.20	2
	3	367	0.10	0.42	0.32	TVM+Γ	(1.2 4.4 2.5 0.4 4.4)	2.56	0	3	1.96	1.78	3
AR	1	387	0.22	0.25	0.32	TIM+Γ	(1.0 4.5 0.5 0.5 3.0)	0.59	0	4	1.94	0.54	4
	2	387	0.27	0.31	0.20	TVM+Γ	(1.2 2.9 0.7 1.8 2.9)	0.71	0	5	0.69	0.55	5
	3	387	0.21	0.31	0.23	TIM+Γ	(1.0 5.4 0.7 0.7 4.4)	1.46	0	6	2.34	1.42	3
GHR	1	284	0.26	0.24	0.33	GTR+Γ	(2.1 3.9 0.9 1.1 2.8)	0.71	0	5	0.90	0.59	5
	2	284	0.31	0.31	0.18	HKY+I+Γ	1.74	1.42	0.28	7	1.43	0.52	4
	3	284	0.21	0.32	0.21	TIM+Γ	(1.0 6.0 0.8 0.8 3.8)	2.69	0	6	2.16	2.45	3
vWF	1	391	0.25	0.28	0.32	TVM+Γ	(1.7 3.4 1.1 1.3 3.4)	0.65	0	5	0.89	0.59	5
	2	391	0.29	0.28	0.17	TrN+Γ+I	(1.0 5.6 1.0 1.0 4.3)	0.81	0.31	8	1.97	0.33	1
	3	391	0.09	0.38	0.40	TVM+Γ	(2.5 9.9 5.6 0.8 9.9)	3.14	0	9	3.02	1.92	3

To assess the stability of the phylogenetic position of *Geogale aurita*, our result was compared, according to both Kishino & Hasegawa (1989) and Shimodaira & Hasegawa (1999) (using RELL bootstrap as well as full optimization methods), to the hypotheses of Olson & Goodman (2003) and Asher & Hofreiter (2006). Furthermore, Ka (i.e. number of nonsynonymous substitutions per nonsynonymous site) and Ks (i.e. number of synonymous substitutions per synonymous site) of pairwise tenrec sequences were calculated using the program CODEML from the PAML package (Yang 1997) in order to assess the molecular divergence between the two *Geogale* GHR sequences and compare it with the level of molecular divergence displayed within the Malagasy tenrec clade.

Molecular dating

We used the Bayesian approach (Thorne et al. 1998) as implemented in the MULTIDIVTIME program package (Thorne & Kishino 2002), which relaxes the molecular clock by allowing continuous autocorrelation of substitution rates among the branches of the phylogenetic tree. The concatenated sequence dataset was partitioned into the same nine categories as for the Bayesian phylogenetic analyses, and branch lengths were calculated under the F84 + Γ model of sequence evolution, which is the most complex model available in MULTIDIVTIME. Each of the described analyses was run twice in order to assess the consistency of the results. The prior for the root was set at 100 Mya, however, analyses with 65 Mya, 80 Mya and 120 Mya as prior age were also performed in order to estimate the impact of the root prior on our results. For each node, we calculated the variance of the estimated ages over all the runs. A maximal variance of $2*10^{-4}$ was found showing that changing the root prior does not influence age estimates. Markov Chain Monte Carlo analyses were run for 1,000,000 generations after a "burn in" of 100,000 generations. The chains were sampled every 100 generations. To assess the influence of a particular partitioning on the dating results, we performed additional analyses using four partitioning schemes: without partitioning, with nine partitions following the results of MODELTEST, with five partitions following the results of ESTBRANCHES using the F84 + Γ model, and with a maximum number of partitions (i.e. twelve). The results of these analyses were close to each other. Notably, all datings for the nodes of interest remained within the 95% credibility intervals of the datings obtained in the analysis using five partitions.

Six well established fossil constraints on divergence times were used: (i) a minimum of 54 and a maximum of 65 Mya for the base of Paenungulata (Gheerbrant et al. 2201); (ii) a minimum of 50 and a maximum of 63 Mya for the split between feliform and caniform Carnivora (McKenna & Bell 1997; Benton 1993); (iii) a minimum of 54 and a maximum of 58 Mya for the split between hippomorph and ceratomorph Perissodactyla (Garland et al. 1993); (iv) a minimum of 55 and a maximum of 65 Mya for the base of Cetartiodactyla (Gatesy & O'Leary 2001); (v) a minimum of 37 Mya for the split between ochotonids and leporids (McKenna & Bell 1997); (vi) a minimum of 60.5 and a maximum of 100.5 Mya for the divergence time between rodents and primates (Benton & Donoghue 2007). To assess the reciprocal

consistency of all calibration points we used the cross-validation method described in (Poux et al. 2005). In this method each calibration point is removed in turn and the remaining calibration points are used to estimate its age. Calibration points, for which the estimated and paleontological dates are not congruent, are considered as inconsistent and are consequently removed from the analyses.

RESULTS AND DISCUSSION

Afrotherian phylogeny

The overall phylogenetic relationships as deduced from the concatenated dataset are consistent with the now broadly accepted branching pattern of the mammalian tree (Murphy et al. 2001) (Fig.1). The superordinal clades Euarchontoglires, Laurasiatheria and Afrotheria are highly supported, and within these clades most bootstrap percentages and posterior probabilities are also high. Afrotheria is now generally accepted as a natural group since molecular studies unanimously support its monophyly, using various methods (Murphy et al. 2001; Nishihara et al. 2005; van Dijk et al. 2001; Robinson et al. 2004; Kellogg et al. 2007). In contrast, until now only few morphological synapomorphies, notably placental morphology (Carter et al. 2006), an increase in number of thoracolumbar vertebrae (Sànchez-Villagra et al. 2007), and testicondy (Werdelin & Nilsonne 1999), appear to support this grouping. Afrotheria are divided into Paenungulata on one hand and the three remaining afrotherian orders (Afrosoricida, Macroscelidea and Tubulidentata) on the other hand. The most probable hypothesis concerning these remaining orders is their grouping within a clade called Afroinsectiphillia (Murphy et al. 2001; Springer et al. 2004; Robinson et al. 2004) within which the internal relationships remain unclear.

Within the paenungulate clade the Tethytheria (elephants + sea cows) are strongly supported by morphological and complete mitochondrial genome data (Novacek 1992; Kjer & Honeycutt 2007). Nuclear genes are ambiguous about this relationship and left the phylogenetic affinities between the three paenungulate orders essentially unresolved (Murphy et al. 2001; Douady & Douzery 2003; Amrine-Madsen et al. 2003; Asher 2007). Our concatenated tree shows for the first time,

based on nuclear genes, strong support for one of the three possible hypotheses: the grouping of Hyracoidea with Sirenia (PP = 0.99 and BP = 89). Bootstrap trees supporting alternative hypotheses exclusively group elephant with hyrax (BP = 11); Tethytheria is never recovered. All four genes independently support this result; the high support for the sea cow + hyrax grouping is therefore expectedly due to the synergy of these non-conflicting informations. To test whether our extensive taxon sampling within Tenrecidae may have improved the phylogenetic accuracy (Zwickl & Hillis 2002; Hillis et al. 2003), all tenrecs but one (*Tenrec ecaudatus*) were removed from a new analysis. The results did not differ much; support for the Sirenia/ Hyracoidea clade dropped negligibly in the concatenated analyses (PP = 0.98 and BP = 86). Interestingly, in a retroposon insertion analysis, Nishihara et al. (2005) found one insertion supporting exclusively the grouping of hyrax with dugong. These authors dismissed the apparent synapomorphous hyrax-sea cow insertion as homoplastic, in favor of the morphological evidence for Tethytheria.

Similarly, the relations between the afroinsectiphillian orders have not yet been clarified, and conclusions vary in different studies. Mitochondrial data give highly inconsistent results (Kjer & Honeycutt 2007; Gibson et al. 2005), while mixed data tend to group golden moles and tenrecs with elephant shrews, together being the sister group of aardvark, with rather strong support (Murphy et al. 2001; Amrine-Madsen et al. 2003; Beck et al. 2006). Our data also support these results, as the Afrosoricida/Macroscelidea clade (= Afroinsectivora) is displayed with high confidence (PP = 1.00 and BP = 93), and Tubulidentata is found to be the sister group of this clade (PP = 1.00 and BP = 95). With a smaller dataset (only one tenrec) the support for the Afrosoricida/Macroscelidea clade slightly increased (PP = 1.00 and BP = 96). Hence, enlarged taxon sampling cannot explain our strong phylogenetic results within the afrotherian clade. All four genes separately displayed Afroinsectiphillia either as paraphyletic or weakly supported therefore the present results are not due to gene sampling biases. The retroposon analyses of Nishihara et al. (2005) proposed the grouping of golden moles, tenrecs and aardvark, to the exclusion of elephant shrews, on the basis of two shared retrotransposons.

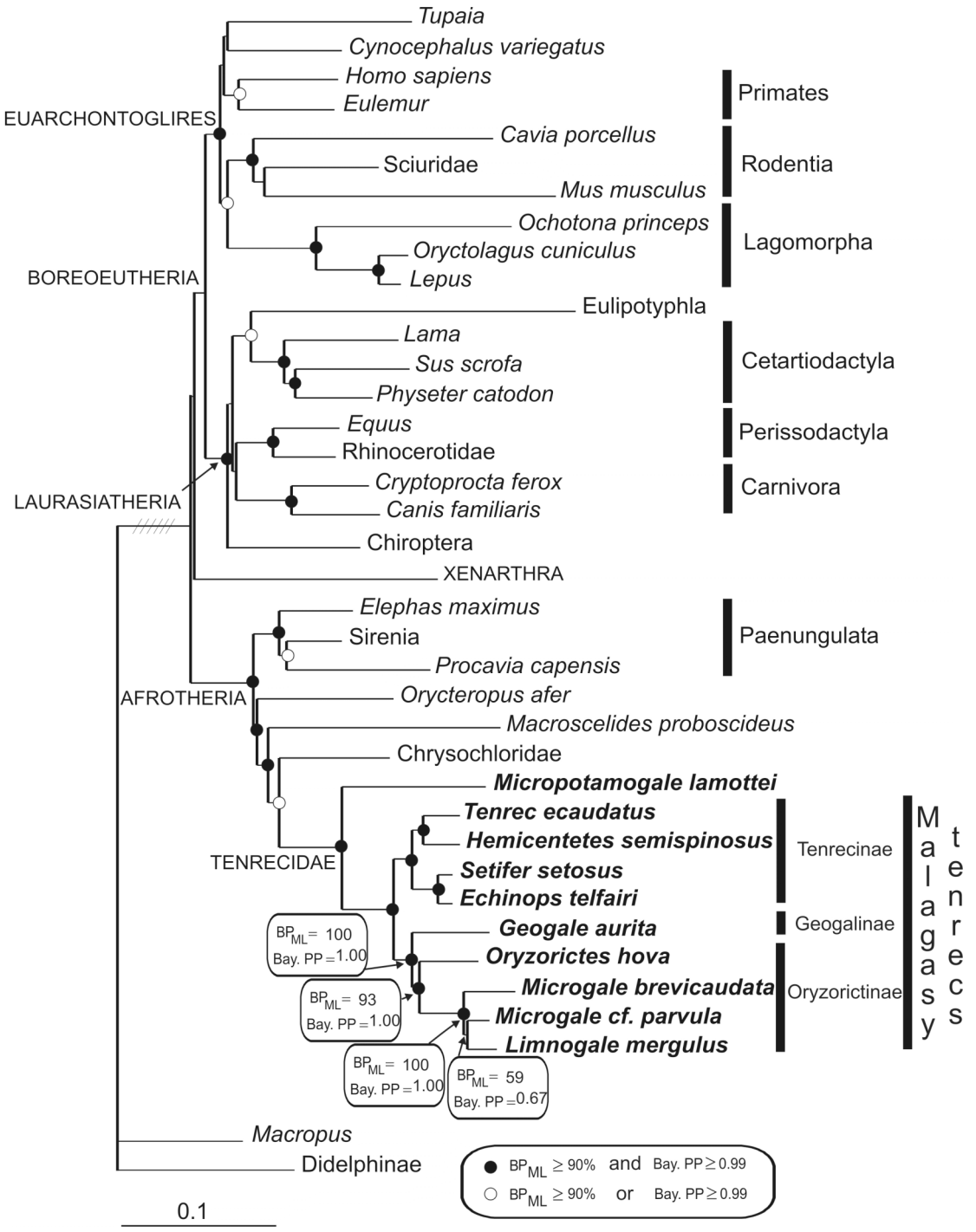

Figure 1: Phylogenetic tree as inferred by maximum likelihood analysis of the concatenated 4,287-bp dataset. Phylogenetic relationships of the investigated mammalian species were reconstructed using ADRA2B, AR, vWF and GHR sequences. Bayesian analyses result in an identical topology. Nodes receiving high support (BP ≥ 90% and PP ≥ 0.99) are marked with filled circles; open circles indicate that nodes received such high support with only one phylogenetic method (either BP or PP). Although the overall phylogenetic relationships as deduced from the present tree are consistent with the broadly accepted branching pattern of the mammalian tree (Murphy et al. 2001), the phylogenetic position of the Eulipotyphla, displaying a high PP node support value, deviates from this consensus. The length of the branch connecting eutherians to the marsupial outgroup was reduced six times. Taxa not indicated by species name are represented by different species in the concatenated dataset, and the higher taxonomic unit is indicated (Table 1).

Phylogenetic position of *Geogale aurita*

The large-eared tenrec (*G. aurita*) has been included until now in only two molecular studies, by Olson & Goodman (2003) and by Asher & Hofreiter (2006). These two studies found two different results concerning its phylogenetic position. The first study, comprising three mitochondrial genes (ND2, 12s rRNA and tRNAvaline) and one nuclear marker (vWF exon 28), displayed, in a parsimony framework, the large-eared tenrec as the most basal of all Malagasy tenrecs. This result was not influenced by the inclusion of morphological characters in the analyses. Asher & Hofreiter (2006), using exon 10 of the GHR gene and morphological data, found *Geogale* nested within the Oryzorictinae, as sister group of the *Microgale*/*Limnogale* clade.

In the present study we also sequenced GHR exon 10 and vWF exon 28, and in addition the intronless gene for ADRA2B and the first exon of AR. For all genes separately the results were congruent in placing *Geogale* as sister group of the Oryzorictinae (Fig.2), although not always strongly supported: ADRA2B: PP = 1.00, BP = 96; AR: PP = 0.77, BP = 86; GHR: PP = 0.64, BP = 59; vWF: PP = 0.93, BP = 61. Concatenation of the four genes led to a stronger support for this node: PP = 1.00 and BP = 93 (Fig.1). The position of *G. aurita* as sister group of the Oryzorictinae (*Oryzorictes*, *Limnogale*, *Microgale*) seems thus strongly supported. However, the KH- and SH-tests (Table 3) did not completely confirm the strength of our results, showing that placing *Geogale* as the most basal Malagasy taxon (Olson & Goodman's hypothesis) was indeed significantly worse than our best tree, but placing it within the Oryzorictinae (Asher & Hofreiter's hypothesis) did not significantly change the likelihood of the topology.

The differences with the results of Olson & Goodman (2003) probably stem from the fact that we did not use the same phylogenetic methods and datasets, even though one of our markers was in common (vWF exon 28). However, their vWF (exon 28) sequences are not yet available in public sources like GenBank to be compared with ours. The different position of *Geogale* in the tree of Asher & Hofreiter (2006) is more difficult to explain. Remarkably, their *Geogale* GHR sequence (Acc. Nr.: DQ202287) displays 18 differences with ours (10 synonymous and 8 non-synonymous substitutions). No mutations leading to unusual amino acid changes that might indicate sequencing errors could be detected. To try and explain

the different *Geogale* GHR sequences we calculated Ka and Ks for each sequence pair of Malagasy tenrecs. The results showed that the sequence divergence between the two *Geogale* specimens was greater than between some of the other tenrec genera, like *Echinops*/*Setifer* and *Limnogale*/ *Microgale* (Table 4). Moreover, the new *Geogale* sequence from this study was slightly more divergent in most comparisons than the one from the database (Table 4). This genetic diversity within *Geogale* could reflect that this genus might contain in fact more than one species. It may also be mentioned that the museum specimen used by Asher & Hofreiter (2006) was collected at the southwest coast of the island (Lamboharana, voucher number MCZ 45044), whereas our specimen (voucher number MVZ mammal # 220648) was sampled in the central west in the Menabe area. Considering photos of living *Geogale* available to us from the south-west (by W. R. Branch) and the central-west (by R. Nincheri and ourselves), the central western specimens appear to have a less golden colored fur and in general a more gracile habitus, but it is unclear whether this may reflect a difference between coloration of adults versus subadults. Clearly, a detailed taxonomic study is needed to confirm whether these differences are constant and the populations may represent two distinct species. Furthermore, a single record of *Geogale* exists also from the east coast near Fenoarivo. This specimen has been described as subspecies *Geogale aurita orientalis* by Grandidier & Petit (1930), but the status of this taxon has remained obscure. It may be a candidate nomen to be elevated to species rank if *Geogale aurita* is demonstrated to consist of more than one species.

Table 3: Results of the Shimodaira-Hasegawa test. RELL and full option test give the same results. The Kishino-Hasegawa test applied to the following hypotheses leads to the same conclusions. Performing the tests including only the Afrotherian species in the analyses does not change the results either.

Trees	Phylogenetic hypothesis	-ln L	Δ –ln L	P
This study	*Geogale* sister group of Oryzorictinae	54619.52	best	
Asher and Hofreiter (2006)	*Geogale* nested within the Oryzorictinae	54632.08	12.56	P = 0.287
Olson and Goodman (2003)	*Geogale* sister group of all other Malagasy tenrecs	54677.72	58.20	P < 0.001

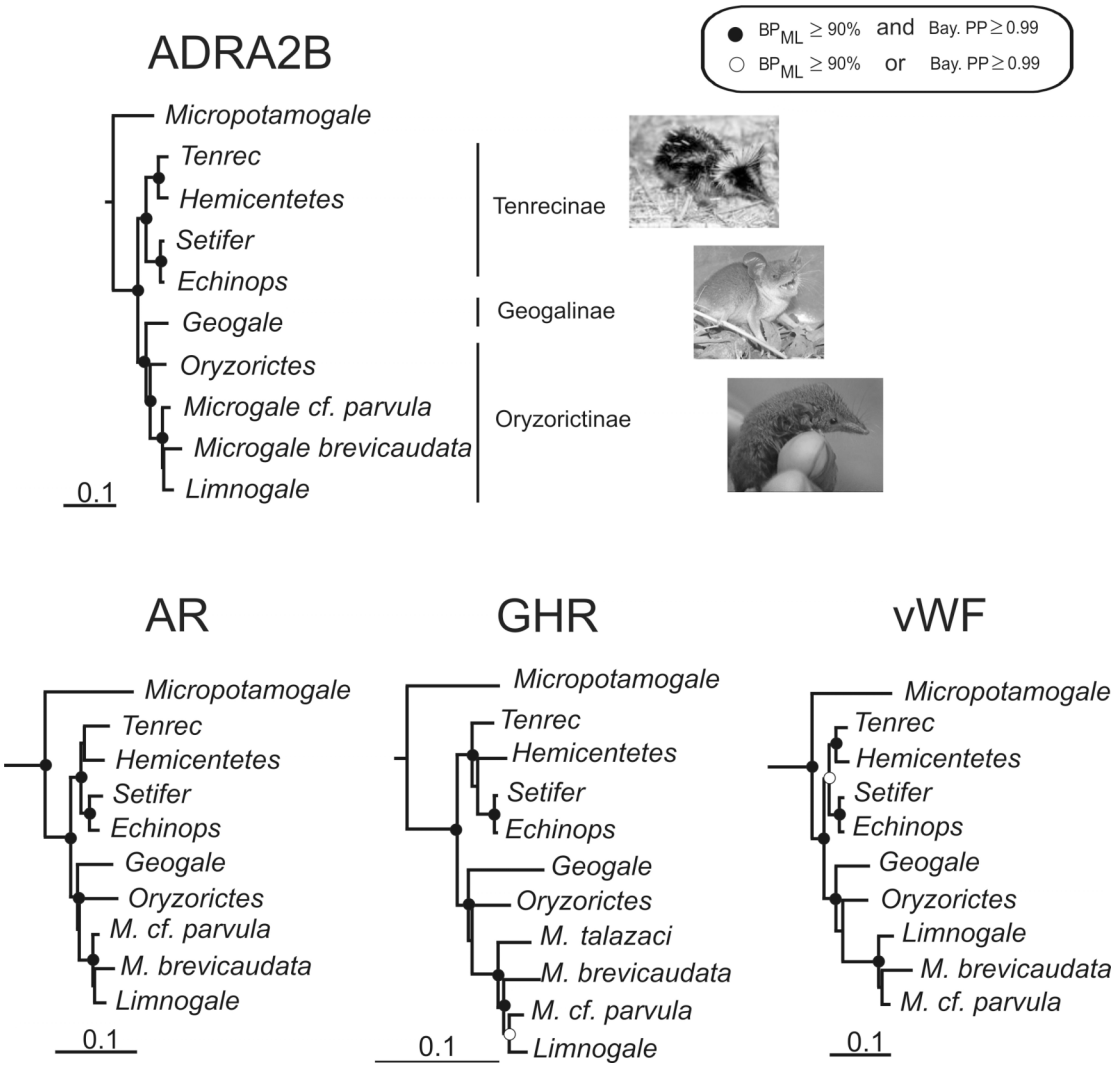

Figure 2: Phylogenetic relationships of tenrecs as inferred by maximum likelihood analysis of the four separate datasets. DNA matrix lengths were 1,101 bp for ADRA2B, 1,161 bp for AR, 852 bp for GHR and 1,173 bp for vWF. Bayesian analyses result in identical topologies. Nodes receiving high support (BP ≥ 90% and PP ≥ 0.99) are marked with filled circles; open circles indicate that nodes received a high support with only one phylogenetic method (either BP or PP). *M. talazaci* sequences were only available for GHR.

Further phylogenetic analyses of the GHR dataset, including both *Geogale* sequences or removing all segregating sites between the two sequences, led to the same result as obtained by Asher & Hofreiter (2006), i.e. *Geogale* nested within the Oryzorictinae. The phylogenetic position of *Geogale* as sister group of Oryzorictinae was only obtained when our sequence alone was used. However, both *Geogale* sequences always grouped together, confirming the identity of our sequence. These results, in combination with the fact that the Oryzorictinae/Geogalinae clade radiated very fast, might make it difficult to reach a final consensus on the evolution of *Geogale*.

Table 4: Ks and Ka calculated for each pair of Malagasy tenrec GHR sequences. Ks (i.e., number of synonymous substitutions per synonymous site) are given in the lower left part of the table and Ka (i.e., number of nonsynonymous substitutions per nonsynonymous site) in the upper right part. The divergence between the two *Geogale* GHR sequences (underligned) is greater than or equal to that between some other tenrec species (bold). *Geogale* A is the sequence from the database (Acc. Nr.: DQ202287), *Geogale* B is our sequence. *Hemicent.* stands for *Hemicentetes*, and *M. brevi.* for *Microgale brevicaudata*.

	Tenrec	*Setifer*	*Hemicent.*	*Echinops*	*Oryzorictes*	*Geogale* A	*Geogale* B	*M.talazaci*	*M.brevi.*	*M.cf.parvula*	*Limnogale*
Tenrec		**0.02**	0.03	**0.02**	0.04	0.06	0.06	0.06	0.07	0.06	0.06
Setifer	0.09		0.03	**0.00**	0.03	0.07	0.07	0.06	0.07	0.07	0.06
Hemicent.	0.08	0.11		0.03	0.04	0.07	0.07	0.07	0.07	0.06	0.07
Echinops	0.09	**0.02**	0.11		0.03	0.06	0.06	0.06	0.07	0.06	0.06
Oryzorictes	0.13	0.19	0.17	0.19		0.05	0.06	0.05	0.06	0.05	0.05
Geogale A	0.14	0.19	0.16	0.19	0.15		**0.02**	0.07	0.07	0.06	0.06
Geogale B	0.16	0.21	0.19	0.21	0.17	**0.05**		0.07	0.08	0.07	0.07
M.talazaci	0.15	0.22	0.21	0.22	0.20	0.19	0.20		0.04	0.03	0.03
M.brevi.	0.17	0.21	0.19	0.22	0.21	0.21	0.24	0.15		0.03	0.04
M.cf.parvula	0.14	0.17	0.14	0.17	0.17	0.17	0.20	0.09	0.09		**0.02**
Limnogale	0.14	0.17	0.14	0.17	0.18	0.20	0.22	0.12	0.10	**0.03**	

From a morphological point of view the phylogenetic relation between *Geogale* and the Oryzorictinae has never been clear. Although most studies gave unresolved results (Asher 1999; Olson 1999 *in* Olson & Goodman 2003, Asher & Hofreiter 2006), two were concordant with ours (Eisenberg 1981, Salton & Szalay 2004), while none has ever argued that *Geogale* was either the sister group of all Malagasy tenrecs or the sister group of the *Limnogale*/*Microgale* clade. Salton and Szalay (2004) reached the conclusion that the tarsal morphology of *Geogale* warrants its status as a separate subfamily, and suggested its closer affiliation with Oryzorictinae than with Tenrecinae.

Three genera of fossil tenrecids – *Erythrozootes*, *Protenrec* and *Parageogale* – from the Kenyan and Namibian Miocene (16–24 Mya; Million years ago) have been discovered until now (Butler 1984; McKenna & Bell 1997; Mein & Pickford 2003). As *Parageogale* is thought to be the sister group of the extant *Geogale aurita* (McKenna & Bell 1997), these data would suggest a more complex dispersal history than the "one time dispersal event" deduced from the monophyly of Malagasy tenrecs. Asher & Hofreiter (2006) were the first to include these three fossil tenrecids in a phylogenetic framework. Their result confirmed the position of the Kenyan fossils as *Geogale*'s closest relatives. However, alternative hypothesis (e.g. monophyly of the Malagasy tenrecs) could not be ruled out indicating the uncertainty of the *Parageogale*/*Geogale* affinity. Recent studies have argued that the sweepstakes dispersal model (dispersal with small and random probability of success) from Africa to Madagascar suffers from many inconveniences, among which the fact that prevailing winds and currents between Africa and Madagascar would be much more likely to favor transports from the island to the African continent, rather than the reverse route (Masters et al. 2006; Stankiewicz et al. 2006). Therefore, if a second dispersal event ever occurred it was most probably from Madagascar to Africa. Olson & Goodman (2003) suggested a basal position of *Geogale* among Malagasy tenrecs and argued that, if true, this would only imply a minimum of two dispersal events, whereas any other scenario would require at least three. However, a back dispersal of *Parageogale* from Madagascar to Africa would only assume a second dispersal event, independent of the phylogenetic position of *Geogale*.

Phylogenetic position of *Limnogale mergulus*

Due to its semi-aquatic life style, shared with the African Potamogalinae, the determination of the phylogenetic relationship of *Limnogale*, the web-footed tenrec, has led to controversies. Its specialized morphological features brought some authors to the conclusion that *Limnogale* was either sister group of the Potamogalinae (Asher 1999) or sister group of all other Malagasy tenrecs (Eisenberg 1981), the semi-aquatic behavior then being seen as an ancestral state and a key element to facilitate over-water dispersal. In contrast, other morphological studies challenged this view by affirming that *Limnogale* had closer relationships to the shrew tenrecs (*Microgale*), and that the semi-aquatic behavior was an example of convergence acquired twice during tenrec evolution (Guth et al. 1959 *in* Olson & Goodman 2003; Olson 1999 *in* Olson & Goodman 2003). This strong affinity between *Limnogale* and *Microgale* has recently also been supported by a study of hind limb muscles (Endo et al. 2006). These authors argue that *Limnogale* may have been derived from a *Microgale*-like terrestrial ancestor. Molecular studies have now confirmed this last hypothesis (Poux et al. 2005; Olson & Goodman 2003; Asher & Hofreiter 2006). Supporting the hypothesis of Olson & Goodman (2003), our study shows that the semiaquatic *Limnogale* is actually nested within the shrew tenrec genus and not a sister clade of it (Fig.1), now with more elaborate analyses and strong support from four nuclear genes.

The phylogenetic supports displayed in the present study are quite low, even with the concatenated dataset (PP = 0.67 and BP = 59), probably due to the fact that the *Microgale/ Limnogale* clade may have radiated very fast (Fig.1). Only one gene, GHR, presents a high PP of 0.99 for the cluster of *Microgale cf. parvula*/*Limnogale mergulus* (Fig.2). The sequencing of more shrew tenrec species (a total of 21 species has been recorded (Bronner & Jenkins 2005; Olson et al. 2004; Goodman et al. 2006; Goodman & Jenkins 2007) might help to resolve this issue, and subsequently to understand the morphological evolution of the aquatic specialization of the web-footed tenrec.

Tenrec diversification timing

Only three studies have previously assessed the timing of tenrec diversification, mainly to understand their colonization pattern (Douady et al. 2002; Douady & Douzery 2003; Poux et al. 2005); none comprised a taxon sampling broad enough to delineate the successive tenrec speciation events. The study by Douady et al. (2002) was based on a linearized tree method and suggests an early diversification of Tenrecs as compared to the other studies (for the present study see Fig.3), which are based on Bayesian methods and partially overlapping gene sampling (Table 5). Consequently, the results of the latter three studies are, as can be expected, rather similar. The present study, with the broadest taxon and gene sampling, estimates the tenrecs/golden mole split at 69 ± 4 Mya, followed by the divergence between African and Malagasy tenrecs at 47 ± 4 Mya. The Malagasy tenrec radiation began 29 ± 3 Mya, and several diversification events spread over time gave rise to the totality of Malagasy tenrec genera around between 20 ± 1 Mya and 7 ± 1 Mya (Table 5 and Fig.3). These datings are slightly older than previously calculated. The only gene difference between this study and Poux et al. (2005) is the inclusion of the GHR gene. Removing it from the calculations led to dates even a little older and with wider confidence intervals (Table 5).

Because the GHR influence on the dating was very small, the difference in taxon sampling between the two studies might be responsible for the different outcomes (Linder et al. 2005). In the present study carnivores and primates were less extensively sampled, whereas Afrosoricida were better represented than in Poux et al. (2005). We therefore compared for these three clades the age inferences in Poux et al. (2005) and in the present study, with or without GHR (Table 6). The conclusion is that the age of a given node tends to become older when the taxon sampling around this node (or descending from it) increases. This phenomenon has already been described by Yoder & Yang (2004) when assessing the timing of evolution of mouse lemurs. They suspected that these incongruences were due to the model used (Thorne & Kishino 2002), which breaks down the path from a tip of the tree to the root (or ancestral node) into identically distributed segments. Such a prior would tend to push divergence time within the clade under study towards unrealistically old ages. Comparing the priors of divergence times between both large and small datasets, they reached the conclusion that the too old priors of the larger dataset had influenced the

posterior estimates, which became older as well. This also is the pattern we can see comparing the priors of Poux et al. (2005) with the ones of the present study (dataset without GHR). In both studies the time estimate differences were not dramatic, but they could have a problematic effect for studies requiring more precise estimates.

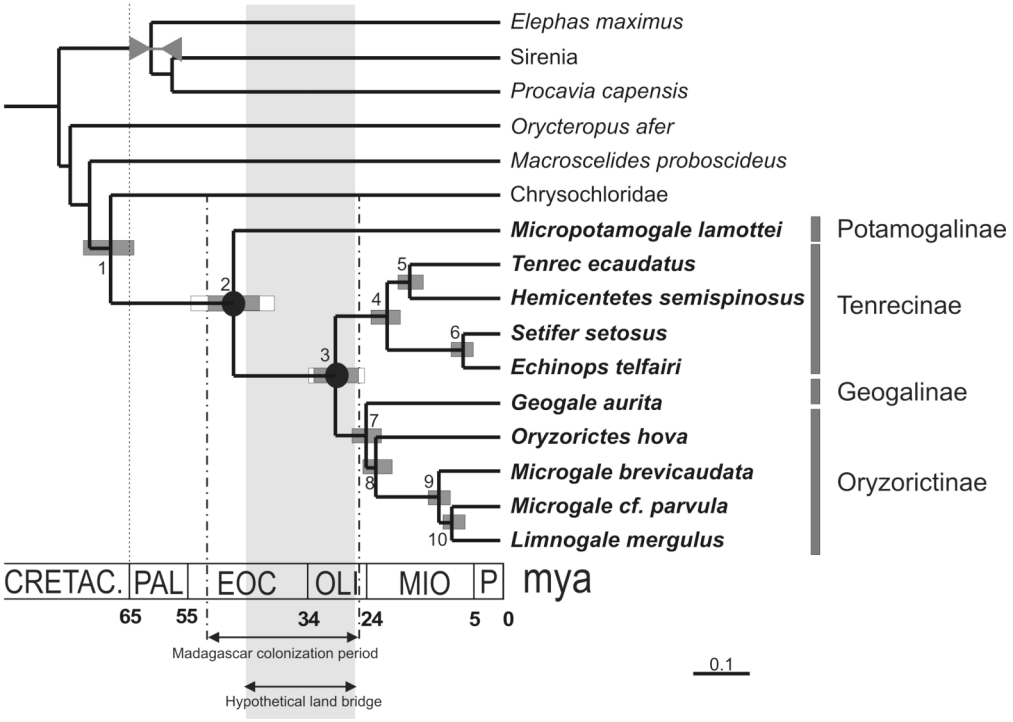

Figure 3: Timing of tenrec speciation events and Madagascar colonization. Tree topology as in Fig.1. Divergence times were estimated from the concatenated dataset by a Bayesian relaxed molecular clock method, with six time constraints from fossil calibrations (see Material and Methods). One of them, the paenungulate radiation is represented on the chronogram. Black circles indicate the divergence from the non-Malagasy sister group (node 2) and the initial divergence of Malagasy tenrecs (node 3). Standard deviations are indicated by grey bars, and 95% credibility intervals by open bars. The period of a putative land bridge between Madagascar and Africa at 45–26 Mya (McCall 1997) is shaded.

The influence of data partitioning was tested as well. The ages of the nodes in the phylogenetic tree increased with the number of partitions (Fig.4A), and the smallest standard deviations (and therefore confidence intervals) were reached for the less extreme numbers of partitions (Fig.4B). However, for the present study, differences in taxon sampling or partitioning did not affect our conclusions, as the various analyses displayed fairly similar results, showing reciprocal overlaps. This is to our knowledge the first time that the influence of data partitioning on dating results has been empirically pointed out. More investigations are needed to generalize and clearly understand the underlying causes of this result. One might however suppose

that the differences between the various partitions could increase with the number of genes included in an analysis. Consequently, these results show that it is important, in order to calculate datings as accurately as possible, to select the right manner of partitioning the data: too few or too many partitions might lead to biased results.

Table 5. Comparison of estimated Malagasy tenrec divergence times (in Mya). Node numbered as in Fig.3. SD: standard deviations; CI: credibility intervals; – Nodes not present in the study.

Clade and node number	Douady et al. (2002)[1]	Douady & Douzery (2003)[2]		Poux et al. (2005)[3]		This study 9 partitions		This study without GHR	
	age	age±SD	95% CI	age±SD	95% CI	age±SD	95% CI	age±SD	95% CI
Tenrecidae/ Chrysochloridae, 1	–	63 ± 5	53-72	67 ± 5	58-76	69 ± 4	61-77	71 ± 4	62-80
Malagasy tenrecs/ Potamogalinae, 2	51-55	43 ± 5	34-52	42 ± 4	34-50	47 ± 4	40-55	45 ± 4	37-54
Malagasy tenrec radiation, 3	37	–	–	25 ± 3	20-32	29 ± 3	24-35	30 ± 3	24-37
Tenrecinae radiation, 4	18-44	16 ± 3	11-22	18 ± 2	13-23	20 ± 2	16-25	21 ± 3	16-26
Tenrec/ *Hemicentetes* split, 5	–	–	–	13 ± 2	10-18	16 ± 2	12-21	15 ± 2	11-20
Setifer/ *Echinops* split, 6	–	–	–	6 ± 1	4-9	7 ± 1	4-9	8 ± 2	5-11
Geogalinae/ Oryzorictinae split, 7	–	–	–	–	–	24 ± 3	19-30	24 ± 3	19-31
Oryzorictinae radiation, 8	–	–	–	19 ± 3	14-25	22 ± 3	17-28	22 ± 3	17-28
Microgale radiation, 9	–	–	–	–	–	11 ± 2	8-15	11 ± 2	7-15
Microgale/ *Limnogale* split, 10	–	–	–	–	–	9 ± 1	6-12	9 ± 2	6-13

[a]Age estimated from vWF, 12s and 16s
[b]Age estimated from vWF, ADRA2B, BRCA1
[c]Age estimated from vWF, ADRA2B, AR

Table 6. Posterior estimates of divergence times (Mya ± standard deviation) inferred from the concatenated datasets. Bayesian relaxed molecular clock method was used. Ages in bold indicate the study in which the corresponding order was more extensively sampled. The result shows that increasing the sampling size pushes the ages towards older estimates. In this analysis rodents could not be taken into account because of sampling incongruence between the two studies.

Radiation	Calibration time frame (Mya)[a]	Poux et al. (2005)[b]	This study without GHR[b]	This study[b]
Primates	60-90	**78.9 ± 4.5**	73.6 ± 4.6	75.7 ± 4.3
Carnivora	50-63	**55.6 ± 3.1**	54.7 ± 3.0	53.3 ± 2.4
Afroinsectiphillia	none	73.7 ± 4.0	**77.3 ± 3.9**	**76.4 ± 3.6**

[a]Paleontological time constraints used as calibrations.
[b]The results of MODELTEST were used to define the partitioning; the three studies are therefore directly comparable.

To exclude the possibility that individual calibration constraints may bias our dating analyses, we repeated them after removing each calibration point in turn following (Poux et al. 2005). Hereby we could check whether the excluded calibration constraint was accurately estimated by the remaining ones. All datings remained

highly congruent when any of the six calibration points was removed. The average percentage difference between the main analysis and the ones with only 5 constrained nodes ranges between 0.1 and 0.8 percent. Only the paenungulate calibration seems to have a somewhat larger impact on the dating as its removal from the analysis increases the estimated node age by 4.8 percent. This influence is however too slight to have an impact on our conclusions. Moreover, the calibrations were reciprocally compatible: the remaining five calibrations always recovered a posterior estimate (± SD) for the excluded node within the time window independently obtained from the corresponding fossil evidence.

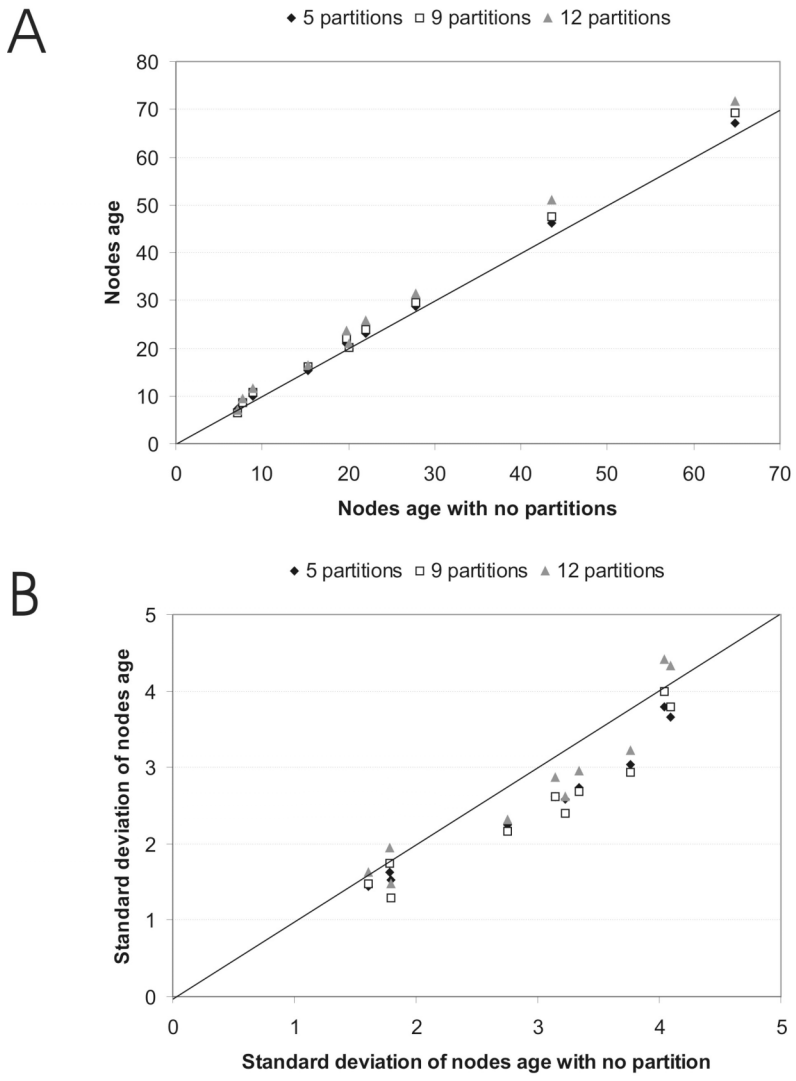

Figure 4: Congruence of divergence time estimates (A) and their associated standard deviations (SD) (B) when calculated with different partition types. The X-axis represents the estimates without partitioning and the Y-axis the ones with 5, 9 or 12 partitions (see Methods). The age estimates increase with the number of partitions (A) and the SDs are larger for extreme numbers of partitions (none and 12 partitions) (B). For clarity purpose only the age estimates relative to tenrecs are displayed in these graphs; however the estimated ages and SDs in the rest of the tree give the same results.

Since *Geogale* has been hypothesized by Olson & Goodman (2003) to be the first Malagasy tenrec genus to have diverged, its absence from Poux et al. (2005) was a problem for drawing final conclusions about tenrec colonization timing. It now appears that *Geogale* is nested within the Malagasy tenrec clade, and therefore plays no role when estimating the period of colonization. Consequently, the window of colonization of Madagascar by tenrecs could not be narrowed. As previously concluded in Poux et al. (2005), the tenrec colonization time completely overlaps with the hypothetical time of existence of a land bridge crossing the Mozambican channel (26–45 Mya; Mc Call 1997) (Fig.3), which however is highly controversial (Krause 2003).

Adaptive radiation often occurs when a species is introduced into a new environment, such as an island. One might therefore expect that the majority of the diversification events within the Malagasy tenrecs would have occurred soon after colonization. However, no such pattern of a diversification burst can be seen at the root of the Malagasy tenrecs, and speciation events seem to be spread through time (Fig.3). This could result from two possible scenarios: either Malagasy tenrecs may actually have experienced a fast adaptive radiation, but most of the resulting taxa are now extinct, or all genera appeared indeed at different periods as a result of a slower speciation rate than expected in case of adaptive radiations. Morphologically, one might speak about adaptive radiation of Malagasy tenrecs, but these morphological adaptations do not seem to have developed within a short time span just after the colonization of Madagascar. The most striking example is the semi-aquatic specialization of the genus *Limnogale*, which dates at most from 11 Mya, i.e. 20–38 My after the colonization of the island. The genus *Microgale* is by far the most speciose amongst tenrecs, being represented by 21 species (Bronner & Jenkins 2005; Olson et al. 2004; Goodman et al. 2006; Goodman & Jenkins 2007), while the remaining genera may not comprise more than one species. The acceleration of the molecular evolutionary rates on the internal branches leading to and within this genus (calculated with MULTIDIVTIME on the tree presented in Fig.1), associated with both a poor phylogenetic resolution between the few *Microgale* species (Figs 1 and 2) and its recency among the tenrec genera (Fig.3 and Table 1), suggests that there has been a fast radiation around 11 Mya that gave rise to the current diversity of *Microgale*. It is interesting to note that the other two endemic mammalian Malagasy genera for which radiation times have been assessed apparently diverged around the

same time as *Microgale*: *Eulemur* at 9.7 Mya and *Microcebus* at 8.7–12 Mya (Yoder & Yang 2004). However, not enough data are yet available to confirm this parallel radiation phenomenon.

Even though the colonization of Madagascar by tenrecs might have taken place during the Eocene, the radiation of the extant species started after Madagascar reached its current geographical subtropical location during the early Oligocene (Smith et al. 1994), with warmer climatological conditions probably similar to the actual ones (Wells 2003). The colonization of Madagascar by carnivores and rodents took place at the end or just after the Oligocene, around 20–23.5 Mya for rodents, and 19–26 Mya for carnivores (data taken from Poux et al. 2005 in order to compare results inferred from similar datasets and methods). These dates are quite close to the periods of appearance of extant tenrec genera: the radiation of Tenrecinae and the split between *Tenrec* and *Hemicentetes* occurred 20 ± 2 Mya and 16 ± 2 Mya, respectively; *Geogale* split from the Oryzorictinae 24 ± 3 Mya; and *Oryzorictes* separated from *Microgale* 22 ± 3 Mya. So five out of the seven tenrec genera (*Limnogale* is taken here as a *Microgale*) diverged soon after the colonization of Madagascar by carnivores and rodents. These new colonizations may have altered the ecological conditions, and thereby induced speciation within tenrecs, either by predation pressure (carnivores) or by interspecific niche competition (rodents).

The complete phylogeny of the Malagasy tenrec genera has now been resolved with strong support. These results should lead to a revision of the taxonomy with regard to the genus *Geogale* (if it comprises more than one species) and the *Limnogale*/*Microgale* clade (if this last genus is truly paraphyletic). This solid phylogenetic and dating framework shows that the major morphological specializations of the tenrecs are not the result of fast adaptive radiations just after colonization, but would as well have been affected by ecological changes caused by climatic and/or subsequent colonization events; however, more work is still needed to understand the role of possible biotic interactions on the speciation processes of Malagasy tenrecs.

ACKNOWLEDGMENTS

We are grateful to numerous colleagues who provided samples and useful information, in particular to W.R. Branch, S. M. Goodman, R. Nincheri, J. Patton, and D.R. Vieites. Field work was carried out in collaboration with the Département de Biologie Animale of the University of Antananarivo. We would like to thank the Malagasy authorities for permits. C.P., O.M. and M.V. were supported by grants from the Netherlands Organization of Scientific Research (NWO).

REFERENCES

Amrine-Madsen H, Koepfli KP, Wayne RK, Springer MS (2003). A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol Phylogenet Evol* 28:225-240.

Asher RJ, Hofreiter M (2006). Tenrec phylogeny and the noninvasive extraction of nuclear DNA. *Syst Biol* 55:181-194.

Asher RJ (1999). A morphological basis for assessing the phylogeny of the "Tenrecoidea" (Mammalia, Lipotyphla). *Cladistics* 15:231-252.

Asher RJ (2007). A web-database of mammalian morphology and a reanalysis of placental phylogeny. *BMC Evol Biol* 7:108.

Beck RM, Bininda-Emonds OR, Cardillo M, Liu FG, Purvis A (2006). A higher-level MRP supertree of placental mammals. *BMC Evol Biol* 6:93.

Benton MJ, Donoghue PC (2007). Paleontological evidence to date the tree of life. *Mol Biol Evol* 24:26-53.

Benton MJ (1993). Fossil record 2. London Chapman and Hall.

Bossuyt F, Milinkovitch MC (2001). Amphibians as indicators of Early Tertiary "out-of-India" dispersal of vertebrates. *Science* 292:93-95.

Bronner GN, Jenkins PD (2005). Order Afrosoricida *in* Mammal species of the world, D.E. Wilson and D.M. Reeder, eds. Pp 71-77. Johns Hopkins Press, Baltimore.

Butler PM (1984). Macroscelidea, Insectivora, and Chiroptera from the Miocene of East Africa. *Palaeovertebrata* 14:117-200.

Carter AM, Blankenship TN, Enders AC, Vogel P (2006). The fetal membranes of the otter shrews and a synapomorphy for Afrotheria. *Placenta* 27:258-268.

Douady CJ, Catzeflis F, Kao DJ, Springer MS, Stanhope MJ (2002). Molecular evidence for the monophyly of Tenrecidae (Mammalia) and the timing of the colonization of Madagascar by Malagasy tenrecs. *Mol Phylogenet Evol* 22:357-363.

Douady CJ, Douzery EJ (2003). Molecular estimation of eulipotyphlan divergence times and the evolution of "Insectivora". *Mol Phylogenet Evol* 28:285-296.

Eisenberg JF (1981). The Mammalian Radiations. An Analysis of Trends of Evolution,

Adaptation and Behavior. University of Chicago Press, Chicago.

Emerson GL, Kilpatrick CW, McNiff BE, Ottenwalder J, Allard MW (1999). Phylogenetic relationships of the order Insectivora based on complete 12S rRNA sequences from mitochondria. *Cladistics* 15:221-230.

Endo H, Yonezawa T, Rakotondraparany F, Sasaki M, Hasegawa M (2006). The adaptational strategies of the hindlimb muscles in the Tenrecidae species including the aquatic web-footed tenrec (Limnogale mergulus). *Ann Anat* 188:383-390.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.

Garland TJ, Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. *Syst Biol* 42:265-292.

Gatesy J, O'Leary MA (2001). Deciphering whale origins with molecules and fossils. *Trend Ecol Evol* 16:562-570.

Gheerbrant E, Sudre J, Larochene M, Moumni A (2001). First ascertained African "Condylarth" mammals (primitive ungulates: cf. Bulbulodentata and cf. Phenacodonta) from the earliest Ypresian of the Ouled Abdoun Basin, Morocco. *J Vert Paleont* 21:107-118.

Gibson A, Gowri-Shankar V, Higgs PG, Rattray M (2005). A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Mol Biol Evol* 22:251-264.

Goodman SM, Jenkins P (2007). *Microgale jenkinsae*. 2007 IUCN Red List of Threatened Species 2006 [http://www.iucnredlist.org]. Downloaded on 29 October 2007

Goodman SM, Raxworthy CJ, Maminirina CP, Olson LE (2006). A new species of shrew tenrec (*Microgale jobihely*) from northern Madagascar. *Journal of Zoology* 270:384-398.

Grandidier G, Petit G. (1930). Etude d'un mammifère insectivore malgache. Le *Geogale aurita* Alph. *in* Faune des Colonies Francaises, Milne-Edwards et Alfred Grandidier, eds. Société des Editions, Paris.

Grant PR (1998). Evolution on Islands. Oxford University Press, Oxford.

Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003). Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* 52:124-126.

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.

Kellogg ME, Burkett S, Dennis TR, Stone G, Gray BA, McGuire PM, Zori RT, Stanyon R (2007). Chromosome painting in the manatee supports Afrotheria and Paenungulata. *BMC Evol Biol* 7:6.

Kishino H, Hasegawa M (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29:170-179.

Kjer KM, Honeycutt RL (2007). Site specific rates of mitochondrial genomes and the phylogeny of eutheria. *BMC Evol Biol* 7:8.

Krause DW (2003). Late Cretaceous vertebrates from Madagascar: a window into Gondwanan

biogeography at the end of the age of dinosaurs *in* The Natural History of Madagascar, S.M. Goodman and Benstead J.P. eds. Pp40-47. Chicago University Press, Chicago.

Linder HP, Hardy CR, Rutschmann F (2005). Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Mol Phylogenet Evol* 35:569-582.

Losos JB, Jackman TR, Larson A, Queiroz K, Rodriguez-Schettino L (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279:2115-2118.

Malia MJ Jr, Adkins RM, Allard MW (2002). Molecular support for Afrotheria and the polyphyly of Lipotyphla based on analyses of the growth hormone receptor gene. *Mol Phylogenet Evol* 24:91-101.

Masters JC, de Wit MJ, Asher RJ (2006). Reconciling the origins of Africa, India and Madagascar with vertebrate dispersal scenarios. *Folia Primatol* 77:399-418.

McCall RA (1997). Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc R Soc Lond B Biol Sci* 264:663-665.

McKenna MC, Bell SK (1997). Classification of Mammals Above the Species Level. Columbia University Press, New York.

Mein P, Pickford M (2003). Insectivora from Arrisdrift, a basal Middle Miocene locality in Southern Namibia. *Mem Geol Surv Namibia* 19:143-146.

Mouchaty SK, Gullberg A, Janke A, Arnason U (2000). Phylogenetic position of the tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of Echinops telfairi. *Zool Scr* 29:307-317.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348-2351.

Nagy ZT, Joger U, Wink M, Glaw F, Vences M (2003). Multiple colonization of Madagascar and Socotra by colubrid snakes: evidence from nuclear and mitochondrial gene phylogenies. *Proc R Soc Lond B Biol Sci* 270:2613-2621.

Nishihara H, Satta Y, Nikaido M, Thewissen JG, Stanhope MJ, Okada N (2005). A retroposon analysis of Afrotherian phylogeny. *Mol Biol Evol* 22:1823-1833.

Novacek MJ (1992). Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Syst Biol* 41:58-73.

Olson LE, Goodman SM, Yoder AD (2004). Illumination of cryptic species boundaries in long-tailed shrew tenrecs (Mammalia: Tenrecidae; *Microgale*): New insights into geographic variation and distributional constraints. *Biol J Linn Soc* 83:1-22.

Olson LE, Goodman SM (2003). Phylogeny and biogeography of tenrecs *in* The Natural History of Madagascar, S.M. Goodman and J.P. Benstead, eds. Pp 1235-1242. Chicago University Press, Chicago.

Philippe H (1993). MUST: a computer package of management utilities for sequences and trees. *Nucleic Acids Res* 21:5264-5272.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-818.

Poux C, Madsen O, Marquard E, Vieites DR, de Jong WW, Vences M (2005). Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst Biol* 54:719-730.

Robinson TJ, Fu B, Ferguson-Smith MA, Yang F (2004). Cross-species chromosome painting in the golden mole and elephantshrew: support for the mammalian clades Afrotheria and Afroinsectiphillia but not Afroinsectivora. *Proc Biol Sci* 271:1477-1484.

Salton JA, Szalay FS (2004). The tarsal complex of Afro-Malagasy Tenrecoidea: a search for phylogenetically meaningful characters. *J Mam Evol* 11:73-104.

Sànchez-Villagra MR, Narita Y, Kuratani S (2007). Thoracolombar vertebral number: the first skeletal synapomorphy for afrotherian mammals. *Syst Biodiv* 5:1-7.

Schluter D (2000). The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

Seehausen O (2006). African cichlid fish: a model system in adaptive radiation research. *Proc Biol Sci* 273:1987-1998.

Shimodaira H, Hasegawa M (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114-1116.

Smith AG, Smith DG, Funnel BM (1994). Atlas of Mesozoic and Cenozoic Coastlines. Cambridge.

Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, Amrine HM, Stanhope MJ (1997). Endemic african mammals shake the phylogenetic tree. *Nature* 388:61-64.

Springer MS, Stanhope MJ, Madsen O, de Jong WW (2004). Molecules consolidate the placental mammal tree. *Trends Ecol Evol* 19:430-438.

Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS (1998). Molecular evidence for multiple origins of insectivora and for a new order of endemic African insectivore mammals. *Proc Natl Acad Sci USA* 95:9967-9972.

Stankiewicz J, Thiart C, Masters JC, de Wit MJ (2006). Did lemurs have sweepstake tickets? An exploration of Simpson's model for the colonization of Madagascar by mammals. *J Biogeogr* 33:221-235.

Swofford DL (2003). PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4 (Sinauer Associates, Sunderland, Massachusetts). Sunderland, Massachusetts: Sinauer Associates.

Thorne JL, Kishino H, Painter IS (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15:1647-1657.

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol* 51:689-702.

van Dijk MA, Madsen O, Catzeflis F, Stanhope MJ, de Jong WW, Pagel M (2001). Protein sequence signatures support the African clade of mammals. *Proc Natl Acad Sci USA* 98:188-193.

Vences M, Vieites DR, Glaw F, Brinkmann H, Kosuch J, Veith M, Meyer A (2003). Multiple overseas dispersal in amphibians. *Proc R Soc Lond B Biol Sci* 270:2435-2442.

Waddell PJ, Kishino H, Ota R (2001). A phylogenetic foundation for comparative mammalian genomics. *Genome Inform* 12:141-154.

Wells NA (2003). Some hypotheses on the Mesozoic and Cenozoic Paleoenvironmental history of Madagascar *in* The Natural History of Madagascar, S.M. Goodman and J.P. Benstead, eds. Pp-16-34. Chicago University Press, Chicago.

Werdelin L, Nilsonne A (1999). The evolution of the scrotum and testicular descent in mammals: a phylogenetical view. *J Theor Biol* 196:61-72.

Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555-556.

Yoder AD, Burns MM, Zehr S, Delefosse T, Veron G, Goodman SM, Flynn JJ (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421:734-737.

Yoder AD, Yang Z (2004). Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol Ecol* 13:757-773.

Zwickl DJ, Hillis DM (2002). Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 51:588-598.

Chapter 7

General Discussion and Prospects

This thesis builds on the recently established well-supported molecular tree of placental mammals (Madsen et al. 2001; Murphy et al. 2001 a, b). Finding this tree has been a very important step as this phylogeny can now be used as a basis for many other investigations, for example, to study the morphological evolution of mammals (Springer et al. 2008), to understand the evolution of genes and proteins, and to be able to choose the best animal models for medical research (Springer & Murphy 2007). During the present study we have taken advantage of this knowledge to reconstruct phylogenies and biogeographical scenarios of island colonization. In this last chapter I will comment upon the main results of this thesis and propose further research lines.

MOLECULAR MAMMALIAN PHYLOGENY

Even though the establishment of placental mammalian relationships at the ordinal level, based on molecular sequence analyses, resulted in a phylogenetic tree with strong resolution (see Fig.8, **chapter 1**), it remained important to search for independent confirmations of the relationships from other genetic sources such as rare genomic changes (RGCs; Rokas and Holland 2000). Not only will such independent data give better confidence in the tree, they also may help to convince any sceptical mammalian systematists that the current molecular tree best approaches the true tree. Rare genomic changes are for example insertions or deletions in protein-coding genes, alternative splicing, and short interspersed elements (SINEs). In eutherian systematics such rare genomic changes have extensively been used as phylogenetic markers. Especially SINEs successfully contributed to reconstruct the evolutionary history of placental mammals (Kriegs et al. 2006), to resolve intraordinal relationships within Cetartiodactyla (Nikaido et al. 1999), Primates (Schmitz et al. 2005) and Xenarthra (Möller-Krull et al. 2007), and interordinal relationships within Afrotheria (Nishihara et al. 2005) and Laurasiatheria (Nishihara et al. 2006). Most recently, large-scale retroposon analyses strongly indicated a nearly concomitant divergence of the three basal placental clades, Xenarthra, Afrotheria and Boreoeutheria (comprising Laurasiatheria and Euarchontoglires), during the near-simultaneous divisions of continents leading to isolated Africa, South America and Laurasia, approximately 120 MYA (Nishihara et al. 2009; Churakov et al. 2009). Indels have been found supporting afrotherian (Madsen et al. 2001) and xenarthran (van Dijk et al. 1999) monophyly, and were searched in genome sequence assemblies to unravel basal placental relationships (Murphy et al. 2007). In **chapter 2** of this

thesis we presented our contribution to the confirmation of the monophyly of Euarchontoglires via the presence of indels in two coding genes (**chapter 2**, Fig.1; Poux et al. 2002). In combination with the improvements in modelling sequence evolution, RGCs can help to solve the last remaining ambiguities in the eutherian molecular tree, or may suggest that certain trichotomies are intrinsically unresolvable.

TESTING DATING METHODS

The use of DNA sequences to estimate the timing of evolutionary events is increasingly popular. Based on the central idea that the differences between the DNA sequences of two species are a function of the time since their separation, molecular dating has been used as a method to investigate both patterns and processes of evolution. During the years of this thesis project the improvements in molecular dating methods have been substantial. This is reflected as well in the chapters of this thesis. In **chapter 3** we only used a local molecular clock, whereas in the following chapters the relaxed clock method has been applied, allowing an autocorrelation of the molecular rate between adjacent branches of the tree (Thorne & Kishino 2002). However, up to now even further improvements have been made but were not tested during this work. Drummond et al. (2006) did not find any significant rate autocorrelation among branches, and proposed a method sampling simultaneously phylogenetic trees and times of diversification (Drummond & Rambaut 2007). It would therefore have been interesting to test with our datasets whether the outcomes of the various state-of-the-art methods would have given the same outcomes.

Moreover, significant methodological challenges are inherent to the use of molecular dating, as described in the Introduction (see **chapter 1, B and C**). Although recent studies have addressed the issue of variation among substitution rates, other difficulties persist and especially concern the impact of the calibration procedure, of data partitioning and sampling strategy on the actual results. Some tests already exist to assess the congruency between calibration points, but do not always lead to the same conclusions (unpublished results). The datasets produced during large-scale phylogenetic studies as in this thesis, helped by simulation studies, can be used to understand how the various dating softwares and methods react to data partitioning and sampling strategy, and therefore estimate their reliability. The sampling strategy refers to the number of species and/or individuals being used in the analyses, and data partitioning involves splitting the dataset in smaller subsets known to

follow different evolutionary processes. In **chapter 6** we have shown that both these parameters may lead to erroneous estimates when they are not correctly assessed. However, these results need to be confirmed with larger and more appropriate datasets in order (i) to know whether they can be generalized or not; (ii) to understand how misleading it can be to take wrong parameters, and (iii) to find criteria for making the right choices.

BIOGEOGRAPHY AND ISLAND COLONIZATION

The biotic enigmas of colonization of Southern America and Madagascar have inspired years of speculation about the mechanisms by which biota reached them. All taxa of terrestrial mammals of Madagascar have now been subjected to rigorous molecular phylogenetic analysis, and all groups have had divergence age estimates generated in at least one study. The results of these different analyses are strikingly uniform. For all four groups (carnivores, lemurs, rodents and tenrecs), phylogenetic analysis demonstrated each group to be monophyletic, with its sister group found in Africa (see **chapter 5**). These patterns lead to the conclusion that each Malagasy clade is the product of a single colonization event and a subsequent radiation of neoendemics within the island. This is true for lemurs (Goodman et al. 1994; Yoder 1994, 1997; Yoder et al. 1996; Poux et al. 2005), rodents (Jansa & Weksler 2004; Poux et al. 2005), tenrecs (Douady et al. 2002; Olson & Goodman 2003; Poux et al. 2005), and carnivores (Yoder & Flynn 2003; Yoder et al. 2003; Poux et al. 2005).

There is a consistent finding among the Malagasy divergence age studies that all crown group ages are spread over the Cenozoic (younger than 65 MY) (Yoder et al. 1996; 2003; Yang & Yoder 2003; Yoder & Yang 2004; Poux et al. 2005) meaning that colonizations do not seem to have occurred during a specific period. With the exception of the tenrecs (Poux et al. 2005, 2008) none of the colonizations could have been facilitated by the hypothetical landbridge between Africa and Madagascar (McCall, 1997). A recently published ancient DNA study confirms that the subfossil giant lemurs belong to the lemuriform clade, and are thus descended from the same Cenozoic colonist that produced the extant lemurs (Karanth et al. 2005). Even though terrestrial mammals are poor over-water dispersers, as indicated by their rareness on isolated oceanic islands (Lawlor, 1986), they have been able to cross the Mozambican channel (~ 400 km) to reach Madagascar by “rafting” or island hopping (e.g., Krause et al. 1997). Yoder & Nowak (2006) reviewed a comprehensive sample of phylogenetic studies of Malagasy biota and indeed found an overwhelming

indication of Cenozoic origins for most Malagasy clades. They concluded that most of the present-day biota of Madagascar is comprised of the descendents of Cenozoic dispersers, predominantly with African origins.

However, to really understand whether there has been a major crossing of the Mozambican channel during a specific lapse of time, it would be necessary to compare the timing of arrival of many clades among all the endemic non-flying fauna (or at least vertebrates). This presents an important and difficult work for the following reasons: 1) it requires a large sequencing effort, 2) one should find markers displaying the right molecular evolutionary rate in order to be able to date and compare monophyletic groups that are phylogenetically extremely distant, 3) one needs to find fossils for calibration points that are spread over the phylogenetic tree to avoid biases due to long distance between the fossil calibrations and the nodes to date. But in case all these requirements would be fulfilled, then one could get enough information to test our hypothesis in a statistical framework.

Biogeographic scenarios for the colonization of South America by primates and rodents have been less studied using molecules. However, with few exceptions (Arnason et al. 2000), molecular datings (Schrago et al. 2003; Poux et al. 2006; Schrago 2007) and fossils (Lavocat 1980; Takai and Anaya 1996; Fleagle 2000) are quite congruent about the age of the separation between old world monkeys (Catarrhini) and new world monkeys (Platyrrhini), which happened around 35-40 MYA. The diversification time of the new world monkeys, 17-20 MYA, has also been consistently recovered by many studies (Schneider et al. 1993; Barroso et al. 1997; Poux et al. 2006; Schrago 2007). All the estimates for the origin of new world monkeys are in agreement with the hypothesis of a transatlantic journey from Africa to South America, as suggested by the fossil record and the tectonic plate movements (South America was not connected anymore to any landmass when primates arrived). This is a surprising result as, at that time, the distance between Africa and South America was around 1500 km. The gap between the Platyrrhini/Catarrhini separation (35-40 MYA) and the last common ancestor of living Platyrrhini (17-20 MYA) may be as big as 20 million years. Paleontological and geoclimatological evidence corroborates that the sudden appearance of modern platyrrhine families may be a consequence of environmental changes during the Miocene (23-5 MYA).

Geoclimatologic studies revealed that the globe has suffered severe environmental modifications during the Cenozoic (Zachos et al. 2001). These changes deeply affected mammalian evolution in South America (MacFadden 2006). The situation is more drastic if we consider mammalian groups with a restricted ecological distribution, such as platyrrhines.

It is possible that the sudden diversification of living platyrrhine families, as recovered by the time scale presented here (**chapter 4**, Fig.2**)** is a consequence of climatic variations. In fact, the occurrence of speciation events that gave rise to modern families corresponds to important shifts in temperature levels and tectonic activity (Zachos et al., 2001). Delsuc et al. (2004), who inferred a time scale for xenarthrans using the same MCMC procedure as in **chapter 4**, reported a similar correspondence between geoclimatologic changes and cladogenetic events. Therefore, it seems that primate evolution in the New World follows the general trends found in other mammalian groups.

Unravelling the biogeographic history of South American caviomorph rodents has proved to be more difficult for two major reasons. First, because it is not yet completely clear whether caviomorph rodents originated in Asia or Africa, and second because they colonized South America in a period during which the continent was still connected to Antarctica and indirectly to Australia. Caviomorphs have, indeed, reached South America around 35-45 MYA (Adkins et al. 2003; Hasegawa et al. 2003; Springer et al. 2003; Poux et al. 2006) and diversified soon thereafter (34-37 MYA) (Hasegawa et al. 2003; Poux et al. 2006). More research is still needed to fully understand their evolutionary history.

The main conclusion that can be drawn from the results presented in **chapters 4** and **5** is that over-water dispersal is a fundamental parameter to explain the distribution and the evolution of mammalian biodiversity on oceanic islands and isolated continents.

EVOLUTION ON ISLANDS

The Tenrecidae count 30 species, amongst which 22 belong to the genus *Microgale*, the shrew tenrecs. *Microgale* species are superficially rather similar in external appearance; most are brown or grey-brown in colour, and in many species the tail is not remarkably longer or shorter than the head and body. In this genus, elucidation of species boundaries is still ongoing, with the description of newly discovered species, the resurrection of others from synonymy, and the emerging molecular evidence for multiple "cryptic" species of shrew tenrecs. Therefore, the taxonomy of this group has not yet led to any consensus, and the number of *Microgale* species depends upon the authors (Heim de Balsac 1972, MacPhee 1987, Wilson and Reeder 2005). MacPhee (1987) revised the genus and divided the species he considered as valid into six morphological groups based on body proportions, dental variation, and other characters, but he explicitly made no attempt at phylogenetic

interpretation. Beside the taxonomical issues, the marked diversity of the *Microgale* genus and the recent discovery of several new species have important implications for conservation, especially in view of the past and present forest habitat destruction in Madagascar (Kremen et al. 2008). Although a few species of *Microgale* occupy diverse habitats and seem capable of adapting to impoverished environmental conditions, it is probable that many species are far less adaptable, and indeed there is evidence that apparently rare species or those believed to be confined to particular habitats may be threatened by habitat destruction.

The colonization of a new environment can be followed by an adaptive radiation, defined as a rapid succession of speciation events leading to a high ecological and phenotypic diversity within a lineage. The study of adaptive radiations on islands or in lakes is essential for understanding processes of speciation and diversification. Therefore, knowing the patterns and timing of the successive diversification events within endemic island clades, which, like tenrecs, display a broad ecological and morphological diversity, might help to better understand this phenomenon. Moreover, recent studies (Poux et al. 2008, Poux et al. *in prep.* a) show that the timing of tenrec diversification might be both correlated with (i) coincident speciation events in other mammalian Malagasy clades, and (ii) the colonization of Madagascar by new mammalian orders. Enlarging the number of tenrec DNA sequences will allow us to both define *Microgale* species on a molecular basis and build a reliable phylogeny of the Tenrecidae. Based on these results we may be able to understand the tenrec morphological evolutionary history and to assess the possible climatic and/or biotic influences on tenrec speciation.

MOLECULAR EVOLUTION

In the postgenomic era, complete genome comparisons have been performed to identify gene functions and their links with biological processes. However, the number of completely sequenced mammalian genomes is still limited. The counterpart of these large-scale analyses is to restrict the studies to just one gene or protein, but considerably increase the taxon sampling. A careful comparison of the variable and conserved regions can reveal which are the sites that might be of crucial importance for the functioning of a protein (e.g. Madsen et al. 2002; van Rheede et al. 2003).

One of the markers sequenced for this thesis research was the Androgen Receptor (AR) (**chapters 5** and **6**). This is one of the proteins responsible for human neurological

diseases caused by CAG expansions, resulting in abnormally long poly-glutamine tracts (poly(Q)) in the encoded proteins. In the case of the AR, such repeats, responsible for Kennedy's disease, are located in the N-terminal transactivation domain of the protein. The AR is a transcription factor mediating the action of androgens, and its transcriptional activity is inversely correlated with poly(Q) length. Comparing the AR sequences of the specimens sequenced in **chapters 5** and **6** (including representatives of all 18 eutherian orders) we could analyze the evolutionary processes modeling poly(Q) tracts in the mammalian AR (Poux et al. *in prep.* b). Analysis of the poly(Q) tracts showed that repeats containing non-CAG codons were longer and more variable than pure CAG tracts, and the presence of these new codons within poly(Q) tracts was better explained by codon duplication rather than multiple point mutations. These results are in contradiction with the studies that usually compare only primates or human/mouse/rat sequences. These previous works proposed that interrupting codons hamper the dynamics of the poly(Q) tracks (Pearson et al. 2005), but with a larger taxonomic sampling we could now show that interrupting codons belong to the repetitive units participating in the tract elongation process.

In conclusion, using phylogenetic and sequence comparison methods to study particular orthologous proteins seems to be a good first step to better understand the most important regions of that protein in terms of function and/or evolutionary dynamics. This information can subsequently be used for further, mainly biomedical, investigations.

CONCLUDING REMARK

During this work, it has been very interesting to understand that integrating diverse scientific subjects is necessary to be able to answer or test scientific hypotheses. This has been, and still is, the most challenging and exciting part of my (thesis)work: being able to place my own research in the broadest, more global context of the present state of scientific knowledge. In our case, building biogeographical scenarios has required taking into account scientific research from very different fields of study: phylogenetics, statistics, paleontology, paleoclimatology, geology, molecular evolution, etc… It is their cross-fertilization that allowed us to reach some meaningful conclusions.

REFERENCES

Adkins, R. M., A. H. Walton, and R. L. Honeycutt. (2003). Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol. Phylogenet. Evol.* 26: 409-420.

Arnason, U., A. Gullberg, A. S. Burguete, and A. Janke. (2000). Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Hereditas* 133: 217-228.

Barroso, C. M. L., H. Schneider, M. P. C. Schneider, I. Sampaio, M. L. Harada, J. Czelusniak, and M. Goodman. (1997). Update on the phylogenetic systematics of New World monkeys: further DNA evidence for placing the pygmy marmoset (*Cebuella*) within the genus *Callithrix*. *Int. J. Prim.* 18: 651-674.

Churakov, G., J.O. Kriegs, R. Baertsch, A. Zemann, J. Brosius, and J. Schmitz. (2009). Mosaic retroposon insertion patterns in placental mammals. *Genome Res*. 19: 868-875.

Delsuc, F., S. F. Vizcaino, and E. J. Douzery. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evol. Biol.* 4: 11.

Douady, C. J., F. Catzeflis, D. J. Kao, M. S. Springer, and M. J. Stanhope. (2002). Molecular evidence for the monophyly of tenrecidae (mammalia) and the timing of the colonization of Madagascar by Malagasy Tenrecs. *Mol. Phylogenet. Evol.* 22: 357-363.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4: e88.

Drummond, A. J., and A. Rambaut. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.

Fleagle, J. G. (2000). The Century of the past: one hundred years in the study of primate evolution. *Evol. Anthro.* 9: 87-100.

Goodman, M., W. J. Bailey, K. Hayasaka, M. J. Stanhope, J. Slightom, and J. Czelusniak. (1994). Molecular evidence on primate phylogeny from DNA sequences. *Am. J. Phys. Anthropol.* 94: 3-24.

Hasegawa, M., J. L. Thorne, and H. Kishino. (2003). Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78: 267-283.

Heim de Balsac, H. (1972). Insectivores *in* Biogeography and ecology in Madagascar, R. Battistini, and G. Richard-Vindard, eds. Pp. 629-660.W. Junk, The Hague.

Jansa, S. A., and M. Weksler. (2004). Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. *Mol. Phylogenet. Evol.* 31: 256-276.

Karanth, K. P., T. Delefosse, B. Rakotosamimanana, T. J. Parsons, and A. D. Yoder. (2005). Ancient DNA from giant extinct lemurs confirms single origin of Malagasy primates. *Proc. Natl. Acad. Sci. USA*. 102: 5090-5095.

Krause, D. W., J. H. Hartman, and N. A. Wells. (1997). Late Cretaceous vertebrates from Madagascar. Implications for biotic changes in deep time *in* Natural change and human impact in Madagascar, S. M. G. B. D. Patterson, ed. Pp. 3-43. Smithsonian Institution Press, Washington.

Kremen, C., A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. Louis, Jr., R. A. Nussbaum, C. J. Raxworthy, A. Razafimpahanana, G. E. Schatz, M. Vences, D. R. Vieites, P. C. Wright, and M. L. Zjhra.

(2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science* 320: 222-226.

Kriegs, J. O., G. Churakov, M. Kiefmann, U. Jordan, J. Brosius, and J. Schmitz. (2006). Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4: e91.

Lavocat, R. (1980). The implication of rodent paleontology and biogeography to the geographical sources and origin of platyrrhine primates *in* Evolutionary biology of New World monkeys and continental drift, R.F. Ciochon, and A.B. Chiarelli, eds. Pp. 93-102. Plenum Press, New York.

Lawlor, T. E. (1986). Comparative Biogeography of Mammals on Islands. *Biol. J. Linnean Soc.* 28: 99-125.

Macfadden, B. J. (2006). Extinct mammalian biodiversity of the ancient New World tropics. *Trends Ecol. Evol.* 21: 157-165.

MacPhee, R. D. E. (1987). The shrew tenrecs of Madagascar : systematic revision and Holocene distribution of Microgale (Tenrecidae, Insectivora). *Am. Mus. Novi.* 2889: 1-45.

Madsen, O., M. Scally, C. J. Douady, D. J. Kao, R. W. DeBry, R. Adkins, H. M. Amrine, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409: 610-614.

Madsen, O., D. Willemsen, B. M. Ursing, U. Arnason, and W. W. de Jong. (2002). Molecular evolution of the mammalian alpha 2B adrenergic receptor. *Mol. Biol. Evol.* 19: 2150-2160.

McCall, R. A. (1997). Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc. R. Soc. Lond. B Biol. Sci.* 264: 663-665.

Moller-Krull, M., F. Delsuc, G. Churakov, C. Marker, M. Superina, J. Brosius, E. J. Douzery, and J. Schmitz. (2007). Retroposed elements and their flanking regions resolve the evolutionary history of xenarthran mammals (armadillos, anteaters, and sloths). *Mol. Biol. Evol.* 24: 2573-2582.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. (2001a). Molecular phylogenetics and the origins of placental mammals. *Nature* 409: 614-618.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. (2001b). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294: 2348-2351.

Murphy W.J., T.H. Pringle, T.A. Crider, M.S. Springer, and W. Miller. (2007) Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* 2007 17:413-421.

Nikaido, M., A. Rooney, and N. Okada. (1999). Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA.* 96: 10261-10266.

Nishihara, H., Y. Satta, M. Nikaido, J. G. Thewissen, M. J. Stanhope, and N. Okada. (2005). A retroposon analysis of Afrotherian phylogeny. *Mol. Biol. Evol.* 22: 1823-1833.

Nishihara, H., M. Hasegawa, and N. Okada. (2006). Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl. Acad. Sci. USA.* 103: 9929-9934.

Nishihara H., S. Maruyama, and N. Okada. (2009) Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc. Natl. Acad. Sci. USA.* 106:5235-5240.

Olson, L. E., and S. M. Goodman. (2003). Phylogeny and Biogeography of Tenrecs *in* The Natural History of Madagascar, S. M. Goodman, and J. P. Benstead, eds. Pp. 1235-1242. Chicago University Press, Chicago.

Pearson, C. E., K. N. Edamura, and J. D. Cleary. (2005). Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6: 729-742.

Poux, C., T. van Rheede, O. Madsen, and W. W. de Jong. (2002). Sequence gaps join mice and men: phylogenetic evidence from deletions in two proteins. *Mol. Biol. Evol.* 19: 2035-2037.

Poux, C., O. Madsen, E. Marquard, D. R. Vieites, W. W. de Jong, and M. Vences. (2005). Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst. Biol.* 54:719-730.

Poux, C., P. Chevret, D. Huchon, W. W. de Jong, and E. J. Douzery. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.* 55: 228-244.

Poux, C., O. Madsen, J. Glos, W. W. de Jong, and M. Vences. (2008). Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses. *BMC Evol. Biol.* 8: 102.

Poux, C., O. Madsen, E. Verheyen, and M. Vences. Molecular phylogeny and divergence times of Malagasy shrew tenrecs (*Microgale*). *In prep.*a

Poux, C., O. Madsen, J. A. M. Leunissen, M. M. Alba, and W. W. de Jong. Variability of polyglutamine tracts in the mammalian androgen receptor: interrupting codons participate in the elongation process. *In prep*.b

Rokas, A., and P. W. Holland. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15: 454-459.

Schmitz, J., C. Roos, and H. Zischler. (2005). Primate phylogeny: molecular evidence from retroposons. Cytogenet. *Genome Res.* 108: 26-37.

Schneider, H., M. P. Schneider, I. Sampaio, M. L. Harada, M. Stanhope, J. Czelusniak, and M. Goodman. (1993). Molecular phylogeny of the New World monkeys (Platyrrhini, primates). *Mol. Phylogenet. Evol.* 12: 225-242.

Schrago, C. G., and C. A. Russo. (2003). Timing the origin of new world monkeys. *Mol. Biol. Evol.* 20: 1620-1625.

Schrago, C. G. (2007). On the time scale of New World primate diversification. *Am. J. Phys. Anthropol.* 132: 344-354.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. (2003). Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100: 1056-1061.

Springer, M. S., and W. J. Murphy. (2007). Mammalian evolution and biomedicine: new views from phylogeny. *Biol. Rev. Camb. Philos. Soc.* 82: 375-392.

Springer, M.S., R.W. Meredith, E. Eizirik, E. Teeling, and W.J. Murphy. (2008). Morphology and placental Mammal Phylogeny. *Syst. Biol.* 57: 499-503.

Takai, M., and F. Anaya. (1996). New specimens of the oldest fossil platyrrhine, *Branisella boliviana*, from Salla, Bolivia. *Am. J. Phys. Anthropol.* 99: 301-317.

Thorne, J. L., and H. Kishino. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51: 689-702.

van Dijk, M. A. M., E. Paradis, F. Catzeflis, and W. W. de Jong. (1999). The virtues of gaps: xenarthran (edentate) monophyly supported by a unique deletion in aA-crystallin. *Syst. Biol.* 48: 94-106.

van Rheede, T., M. M. Smolenaars, O. Madsen, and W. W. de Jong. (2003). Molecular evolution of the mammalian prion protein. *Mol. Biol. Evol.* 20: 111-121.

Wilson, D. E., and D. M. Reeder. (2005). Mammal species of the world: a taxonomic and geographic reference. Smithsonian Institution Press, Washington and London.

Yoder, A. D. (1994) Relative position of the Cheirogaleidae in strepsirhine phylogeny: a comparison of morphological and molecular methods and results. *Am. J. Phys. Anthropol.* 94: 25-46.

Yoder, A. D., M. Cartmill, M. Ruvolo, K. Smith, and R. Vilgalys. (1996). Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. USA* 93: 5122-5126.

Yoder, A. D. (1997). Back to the future: a synthesis of strepsirrhine systematics. *Evol. Anthro.* 6: 11-22.

Yoder, A. D., and J. J. Flynn. (2003). Origin of Malagasy Carnivora *in* The Natural History of Madagascar, S. M. Goodman, and J. P. Benstead, eds. Pp. 1253-1256. Chicago University Press, Chicago.

Yoder, A. D., M. M. Burns, S. Zehr, T. Delefosse, G. Veron, S. M. Goodman, and J. J. Flynn. (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421: 734-737.

Yoder, A. D., and Z. Yang. (2004). Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* 13: 757-773.

Yoder, A. D., and M. D. Nowak. (2006). Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annu. Rev. Ecol. Evol. Syst.* 37: 405-431.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292: 686-693.

Chapter 8

Summary / Samenvatting / Résumé

The present work mainly focuses on the biogeographic history of endemic mammalian clades that colonized South America (primates and rodents) and Madagascar (primates, rodents, carnivores and tenrecs) during the Cenozoic period, the last 65.5 million years. Since the acceptance of the continental drift theory, a majority of island endemics has been considered to be the result of vicariant evolution. However, given the relatively recent age of the radiation of placental mammals (beginning around 100 million years ago, Mya) compared to the age of the break-up of Gondwana (between 160 and 30 Mya), the present distribution patterns cannot all be explained by vicariance. Therefore several questions have arisen concerning the number, the tempo and the mode of mammalian migrations. How often did each mammalian order colonize South America and Madagascar? Did they use temporary land bridges, i.e. continental connections, or did they somehow cross large water areas? When did these events happen? Were the colonization events synchronized between orders? Answering these questions will teach us about the dispersal and adaptive abilities of mammals. The biogeographical questions as well as a historical perspective on the knowledge of rates of molecular evolution and a review of molecular dating methods are presented in the Introduction, **chapter 1**.

Molecular phylogenetics first of all requires the choice of adequate molecular markers. Most popular as markers are DNA sequences, but also 'rare genomic changes' (RGCs) are often used because they are less prone to homoplasia. **Chapter 2** underlines the importance in phylogeny of one type of RGC, the indels (insertions/deletions). Thanks to two unique deletions present in two different proteins, SCA1 (ataxin 1) and PRNP (prion protein), we could confirm the close phylogenetic relationship between men and mice, as representatives of the orders Primates and Rodentia, respectively, now grouped in a newly recognized superordinal clade called Euarchontoglires.

Chapter 3 deals with the phylogenetic relationships within the primate order, including the platyrrhines from South America and the lemuriforms from Madagascar. This study helped resolving some ambiguous phylogenetic positions, namely those of the Tarsiidae, and of the Cheirogaleidae and Daubentoniidae from Madagascar, using 1200 bp of the IRBP gene (Interstitial Retinoid-Binding Protein). In this dataset rate heterogeneity between lineages was confirmed, and cladistic events were therefore dated via maximum likelihood local molecular clocks, with six independent fossil calibration points. Using this method, the origin of the Lemuriformes (crown group) was dated at between 39.6 Mya and 40.7 Mya.

In **chapter 4,** the platyrrhine primates and caviomorph rodents of South America were

investigated using a dataset composed of three nuclear genes, IRBP, vWF (von Willebrand Factor) and ADRA2b (ADRenergic Alpha 2b receptor). Platyrrhines as well as caviomorphs appear to be monophyletic, reflecting a unique colonization event for each of them. This time a Bayesian relaxed molecular clock approach was used to better accommodate the rate heterogeneity between genes and lineages. According to this approach, platyrrhine primates colonized South America between 37.0 ± 3.0 Mya and 16.8 ± 2.3 Mya, and caviomorph rodents between 45.4 ± 4.1 Mya and 36.7 ± 3.7 Mya. Due to confidence interval overlap, we could not rule out the possibility of a concomitant arrival of primates and rodents in South America. Primates probably migrated by crossing the Atlantic from Africa around the transition Eocene-Oligocene. The colonization of South America by rodents more likely happened during the Middle Eocene, at a time during which the connections between the Southern continents were tighter, so that migration via terrestrial routes over Antarctica cannot be excluded for them.

Chapter 5 describes basically the same type of study but now applied to the four radiations of endemic terrestrial Malagasy mammals. The phylogenetic reconstructions and the datings were based again on three nuclear genes: ADRA2b, vWF and AR (Androgen Receptor). All radiations displayed the same pattern: the four groups of endemic Malagasy mammals are all monophyletic and from African origin. We estimated that lemurs colonized Madagascar between 60 and 50 Mya, tenrecs between 42 and 25 Mya, carnivorans between 26 and 19 Mya, and rodents between 24 and 20 Mya. These results suggested at least two asynchronous colonization events (lemurs vs. rodents and carnivores). Moreover, the colonization of Madagascar by at least lemurs, rodents and carnivores appears to have occurred by overseas rafting rather than via a land bridge hypothesized to have existed between 45 and 26 Mya. The time of colonization by tenrecs allows the use of this land bridge.

The last research chapter, **chapter 6**, focuses on the poorly known tenrec family, mainly present in Madagascar. Phylogenetic and dating analyses were applied to four nuclear gene datasets: ADRA2b, AR, GHR (Growth Hormone Receptor) and vWF, including all Malagasy tenrec genera. Strongly supported relationships were recovered, allowing us to firmly establish some debated phylogenetic positions at the genus level. Given our datings, the Malagasy tenrecs diversified during their whole evolutionary history, and not just within a short period after the colonization of Madagascar, possibly in response to environmental changes caused by climatic and/or subsequent colonization events.

A discussion about the findings of the present research is given in **Chapter 7**.

Mammals are known to be very poor dispersers over water. However, we have shown that they were able to repeatedly colonize continents or islands far from their continent of origin. The observed present mammalian radiations are surviving lineages, but colonization of South America and Madagascar may have happened more often by extinct lineages. Our results show that different scenarios are possible to describe diversification after colonization: a fast diversification into the main lineages that are still extant (as for caviomorph rodents), a postponed burst of diversification (platyrrhine primates) or a step by step diversification (Malagasy tenrecs). It would be interesting to confront these findings with the fossil record, which is, however, rather poor in South America and non-existent in Madagascar.

Het onderzoek in dit proefschrift richt zich voornamelijk op de biogeografische geschiedenis van endemische zoogdiergroepen die Zuid Amerika (primaten en knaagdieren) en Madagascar (primaten, knaagdieren, carnivoren en tenreks) koloniseerden gedurende de laatste 65,5 miljoen jaar, het Cenozoïcum. Sinds de aanvaarding van de "continental drift" theorie worden de meeste endemische soorten op eilanden en geïsoleerde continenten beschouwd als het resultaat van vicariante evolutie, dat wil zeggen, ontstaan doordat populaties gescheiden raakten door oceanen of andere geografische barrières. Echter, gezien het relatief recente uiteengaan van de verschillende groepen van placentale zoogdieren (vanaf ongeveer 100 miljoen jaar geleden) vergeleken met het uiteenvallen van het Gondwana supercontinent (tussen 160 en 30 miljoen jaar geleden), kunnen niet alle huidige verspreidingspatronen verklaard worden door vicariantie. Dit heeft verschillende vragen opgeroepen over het aantal, het tempo en de wijze van zoogdiermigraties. Hoe vaak koloniseerden de genoemde zoogdierorden Zuid Amerika en Madagascar? Gebruikten ze daarbij tijdelijke landbruggen of lukte het ze om uitgestrekte watervlaktes over te steken? Wanneer gebeurde dat? Vonden de kolonisaties door de verschillende zoogdiergroepen gelijktijdig plaats? Door het beantwoorden van deze vragen kunnen we meer te weten komen over de mogelijkheden tot verspreiding en aanpassing van zoogdieren. De betreffende biogeografische vragen, zowel als een historisch overzicht van onze kennis van moleculaire evolutiesnelheden en moleculaire dateringsmethoden, nodig om deze vragen te beantwoorden, worden gepresenteerd in **Hoofdstuk 1**.

Het maken van moleculaire stambomen vereist op de eerste plaats de keuze van geschikte moleculaire kenmerken. Het vergelijken van overeenkomstige DNA volgordes bij de onderzochte organismen is het meest gebruikelijk, maar ook zogenaamde 'rare genomic changes' (RGCs), dus zelden voorkomende veranderingen in het genoom, worden vaak benut. RGCs hebben het voordeel dat ze minder homoplasie vertonen, d.w.z. minder vaak door parallelle of terugmutatie ontstaan dan het geval is bij enkelvoudige base-veranderingen in het DNA. **Hoofdstuk 2** benadrukt het belang van één zo'n type RGC, inserties en deleties ('indels') in DNA of eiwit, bij het fylogenetisch onderzoek. Dankzij twee unieke deleties aanwezig in twee verschillende eiwitten, SCA1 (ataxine 1) en PRNP (het prion eiwit), konden we de relatief nauwe verwantschap bevestigen tussen mens en muis, als vertegenwoordigers van de ordes Primates en Rodentia. Deze twee ordes worden tegenwoordig gegroepeerd in een nieuwe superorde, genaamd Euarchontoglires.

Hoofdstuk 3 behandelt de fylogenetische verwantschappen binnen de orde Primates, met inbegrip van de Zuid-Amerikaanse apen (Platyrrhini) en de lemuren (Lemuriformes) van

Madagascar. Deze studie heeft er toe bijgedragen om enkele twijfelachtige fylogenetische posities te verhelderen, met name die van de familie Tarsiidae, de spookdiertjes, en van de lemurfamilies Cheirogaleidae en Daubentoniidae van Madagascar. Hierbij werd gebruik gemaakt van 1200 baseparen van het IRBP gen (Interstitial Retinoid-Binding Protein). Binnen deze dataset werd heterogeniteit vastgesteld van de evolutiesnelheid tussen verschillende takken. De datering van de vertakkingen gebeurde daarom met een maximum likelihood methode met lokale moleculaire klokken, i.p.v. een universele klok voor de hele boom. Zes onafhankelijke fossiele calibreringspunten werden gebruikt. Met behulp van deze methode werd de oorsprong van de thans nog levende Lemuriformes gedateerd tussen 39,6 en 40,7 miljoen jaar geleden.

In **hoofdstuk 4** werden de endemische platyrrhine primaten en caviomorfe knaagdieren van Zuid Amerika bestudeerd met behulp van een dataset bestaande uit drie kerngenen, IRBP, vWF (von Willebrand Factor) en ADRA2b (ADRenergische Alpha 2b receptor). In dit geval werd gebruik gemaakt van een Bayesiaanse "relaxed molecular clock" benadering om beter rekening te kunnen houden met de heterogeniteit van de evolutiesnelheden tussen genen en evolutietakken. Zowel de Platyrrhini als de Caviomorpha blijken monofyletisch te zijn, wat er op wijst dat kolonisatie van Zuid Amerika door elk van deze twee groepen slechts één maal heeft plaats gevonden. Volgens deze benadering werd Zuid Amerika tussen 37,0 ± 3,0 en 16,8 ± 2,3 miljoen jaar geleden gekoloniseerd door de platyrrhine primaten en tussen 45,4 ± 4,1 en 36,7 ± 3,7 miljoen jaar geleden door de caviomorfe knaagdieren. Door overlap van de betrouwbaarheidsintervallen konden we de mogelijkheid van een gelijktijdige aankomst van primaten en knaagdieren in Zuid Amerika niet geheel uitsluiten. De primaten migreerden waarschijnlijk vanuit Afrika door de Atlantische oceaan over te steken ten tijde van de overgang van Eoceen naar Oligoceen. De kolonisering van Zuid Amerika door knaagdieren vond waarschijnlijk plaats in het Midden of Late Eoceen, gedurende een periode waarin de zuidelijke continenten nog nauwer met elkaar verbonden waren, zodat migratie via landroutes over Antarctica voor hen niet kan worden uitgesloten.

Hoofdstuk 5 beschrijft een soortgelijk onderzoek, maar nu toegepast op de divergenties van de vier groepen endemische landzoogdieren van Madagascar. De fylogenetische reconstructies en dateringen waren weer gebaseerd op drie kerngenen: ADRA2B, vWF en AR (Androgeen Receptor). Elk van de divergenties toonde hetzelfde patroon: de vier groepen van endemische malagasische zoogdieren zijn alle monofyletisch en van Afrikaanse oorsprong. Wij schatten dat de lemuren Madagascar tussen 60 en 50 miljoen

jaar geleden koloniseerden, de tenreks tussen 42 en 25 miljoen jaar geleden, de carnivoren tussen 26 en 19 miljoen jaar geleden en de knaagdieren tussen 24 en 20 miljoen jaar geleden. Deze resultaten suggereerden tenminste twee niet-synchrone kolonisatieperiodes: één van de lemuren en één van knaagdieren en carnivoren; de kolonisatietijd van de tenrecs kan daar tussenin gelegen hebben of overlappend met die van knaagdieren en carnivoren. Bovendien lijkt de kolonisatie van Madagascar door de lemuren, knaagdieren en carnivoren in ieder geval te hebben plaats gevonden door "rafting" overzee vanuit Afrika en niet via een landbrug waarvan het bestaan tussen 45 en 26 miljoen jaar geleden wel verondersteld is. De gevonden kolonisatietijd voor de tenreks zou het gebruik van deze hypothetische landbrug wel mogelijk hebben kunnen maken.

Hoofdstuk 6, het laatste onderzoekshoofdstuk, richt zich op de weinig bekende tenrekfamilie, die voornamelijk op Madagascar voorkomt. Fylogenetische en dateringsanalyses werden uitgevoerd op datasets van vier kerngenen: ADRA2B, AR, GHR (Groei Hormoon Receptor) en vWF, waarin alle genera van malagasische tenreks vertegenwoordigd waren. Er werden sterk ondersteunde verwantschappen gevonden, waardoor we enkele op genus-niveau omstreden fylogenetische posities met grote zekerheid konden vast stellen. Gegeven onze dateringen vond de diversificatie van de malagasische tenreks plaats gedurende hun hele evolutionaire geschiedenis en niet slechts binnen een korte periode direct na de kolonisatie van Madagascar. Deze stapsgewijze diversificatie was mogelijk een reactie op omgevingsveranderingen veroorzaakt door klimatologische en/of latere kolonisatiegebeurtenissen.

De bevindingen zoals gepresenteerd in dit proefschrift worden bediscussieerd in **Hoofdstuk 7**. Het is bekend dat landzoogdieren een zeer beperkt vermogen hebben om zich over water te verbreiden. Toch hebben we laten zien dat ze in staat zijn geweest om herhaaldelijk continenten of eilanden te koloniseren die ver van hun continent van oorsprong lagen. De onderzochte zoogdiergroepen zijn overlevende fylogenetische takken van hun ordes, maar we kunnen niet uitsluiten dat kolonisering van Zuid Amerika en Madagascar vaker heeft plaats gevonden door uitgestorven takken van de betreffende ordes. Onze resultaten laten zien dat verschillende scenario's mogelijk zijn om de diversificatie binnen een zoogdiergroep na kolonisatie van een eiland of continent te beschrijven: een snelle diversificatie na aankomst leidend tot de belangrijkste nog bestaande takken (zoals bij de caviomorfe knaagdieren), een uitgestelde uitbarsting van diversificatie (platyrrhine primaten) of een stap-voor-stap diversificatie (malagasische tenreks). Het zou interessant zijn om deze bevindingen te vergelijken met fossiele vondsten, die echter tamelijk schaars zijn in Zuid

Amerika en afwezig in Madagascar.

Le travail de recherche présenté ici décrit l'histoire biogéographique de mammifères endémiques d'Amérique du Sud (primates platyrrhiniens et rongeurs caviomorphes) et de Madagascar (lémurs, rongeurs, carnivores et tenrecs) pendant la période du Cénozoïque, période qui a commencé il y a 65,5 millions d'années (Ma). Depuis la découverte de la dérive des continents, la majorité des organismes endémiques insulaires a été considérée comme résultant d'une évolution par vicariance. Cependant, étant donné l'âge relativement récent de la radiation des mammifères placentaires (100 Ma environ) par rapport à la période de morcellement du Gondwana (qui a commencé il y a 160 Ma et qui s'est achevée il y a 30 Ma), le phénomène de vicariance ne peut être invoqué pour expliquer la distribution géographique des espèces étudiées ici. Par conséquent, un certain nombre de questions ont émergé concernant le nombre et le tempo des migrations ainsi que le mode migratoire des mammifères. Combien de fois un ordre de mammifère donné a-t-il colonisé l'Amérique du Sud et/ou Madagascar ? Ont-ils utilisés des ponts continentaux (connexions inter-continentales) ou ont-ils pu traverser de larges étendues maritimes ? Quand ces événements ont-ils eu lieu et ont-ils eu lieu simultanément ou en décalé ? Pouvoir répondre à ces questions peut nous apprendre beaucoup sur les capacités adaptatives des mammifères. Ces questions biogéographiques ainsi qu'une revue des connaissances concernant les taux d'évolution moléculaire et les méthodes de datation moléculaire sont présentées dans l'introduction, **chapitre 1**.

La phylogénie moléculaire requiert en tout premier lieu de choisir des marqueurs moléculaires adaptés. Les marqueurs les plus populaires sont les séquences d'ADN, mais il existe aussi d'autres marqueurs comme les variations génomiques rares qui sont souvent utilisées car elles sont peu soumises à l'homoplasie. Le **chapitre 2** souligne l'importance d'un type de variation génomique rare en phylogénie, les indels (insertions/délétions). Grâce à deux délétions uniques présentes dans deux protéines différentes, SCA1 (Ataxine 1) and PRNP (protéine du prion), nous avons pu confirmer l'affinité phylogénétique qui existe entre l'ordre des primates (par exemple l'homme) et l'ordre des rongeurs (par exemple la souris), maintenant regroupés dans un clade superordinal nommé Euarchontoglires.

Dans le **chapitre 3** nous avons étudié les relations phylogénétiques à l'intérieur de l'ordre des primates, incluant les primates platyrrhiniens d'Amérique du Sud et les Primates lémuriformes de Madagascar. Cette étude a permis de résoudre

certaines positions phylogénétiques jusqu'alors ambiguës : celles des Tarsiidae et celles des Cheirogaleidae et des Daubentoniidae de Madagascar. Pour se faire, 1200 paires de bases du gène IRBP (Protéine Interstitielle se liant au Rétinol) ont été utilisées. Ces données ont permis de confirmer la présence d'un taux d'évolution hétérogène entre les lignées de primates. Par conséquent, les événements de cladogénèse ont été datés par des horloges moléculaires locales grâce à la contribution de six fossiles indépendants utilisés comme points de calibration. En utilisant cette méthode, l'origine des lémuriformes remonte à une époque vieille de 39,6 à 40,7 Ma environ.

Au **chapitre 4** les primates et les rongeurs d'Amérique du Sud ont été étudiés à l'aide d'un jeu de données composé de 3 gènes nucléaires, IRBP, vWF (Facteur von Willebrand) et ADRA2b (récepteur ADRénergique Alpha 2b). Les primates platyrrhiniens et les rongeurs caviomorphes apparaissent monophylétiques dans l'arbre phylogénétique ainsi reconstruit. Ce résultat indique que chaque lignée a colonisé le continent sud américain une seule et unique fois. Lors de cette étude, une horloge moléculaire relaxée a été utilisée de façon à mieux appréhender les hétérogénéités de taux d'évolution entre les gènes et les lignées. Selon cette approche, les primates platyrrhniens ont colonisé l'Amérique du Sud pendant une période comprise entre 37,0 (± 3,0) et 16,8 (± 2,3) Ma alors que les rongeurs caviomorphes semblent avoir colonisé le continent plus tôt pendant une période comprise entre 45,4 (± 4,1) et 36,7 (± 3,7) Ma. Les intervalles de confiance entre ces deux périodes se superposent légèrement, par conséquent nous ne pouvons pas exclure la possibilité d'une arrivée concomitante des primates et des rongeurs en Amérique du Sud. Les primates ont probablement effectué leur migration en traversant l'Océan Atlantique à une période proche de la transition Eocene-Oligocene. La colonisation de l'Amérique du Sud par les rongeurs a plus vraisemblablement eu lieu lors de l'Eocène moyen. Leur migration a pu s'effectuer par voies terrestres via l'Antarctique car, pendant cette période, les terres australes étaient encore connectées entre elles.

Le **chapitre 5** décrit le même type d'étude mais cette fois appliquée aux quatre radiations de mammifères endémiques de Madagascar (les lémurs, les tenrecs, les carnivores et les rongeurs). Les reconstructions phylogénétiques et les datations moléculaires ont de nouveau été réalisées grâce à trois gènes nucléaires : ADRA2b, vWF et AR (Récepteur Androgène). Les résultats pour les 4 groupes de mammifères sont semblables, ils démontrent la monophylie et l'origine Africaine de chacun. Selon

nos estimations, la colonisation de Madagascar par les lémurs a eu lieu durant une période comprise entre 60 et 50 Ma, celle des tenrecs entre 42 et 25 Ma, celle des carnivores entre 26 et 19 MA et enfin celle des rongeurs entre 24 et 20 Ma. Ces résultats suggèrent qu'il y a eu au moins deux épisodes de colonisation : la colonisation de Madagascar par les lémurs étant relativement vieille comparée à celle des carnivores et des rongeurs, ces événements n'ont pas eu lieu pendant la même période. De plus ces trois événements de colonisation ont eu lieu grâce à des radeaux et non grâce à des connexions continentales qui auraient existé durant une période comprise entre 45 et 26 millions d'années. A contrario, les tenrecs pourraient avoir fait usage de cette connexion continentale pour coloniser Madagascar.

Le dernier chapitre de recherche, le **chapitre 6**, porte son intérêt sur une famille de mammifères peu connue, la famille des tenrecs présente essentiellement à Madagascar. Les analyses phylogénétiques et les datations ont été réalisées grâce à un jeu de données composé de 4 gènes, ADRA2b, AR, GHR (Récepteur à l'hormone de croissance) et vWF, et incluant tous les genres de tenrecs malgaches. Le résultat des analyses phylogénétiques permet d'établir clairement des relations phylogénétiques jusqu'alors débattues. Selon le résultat des datations, les tenrecs de Madagascar se sont diversifiés durant toute leur période évolutive, et pas seulement pendant une période restreinte à leur arrivée sur l'île. Les tenrecs se sont diversifiés probablement en réponse à des changements environnementaux liés à des événements climatiques et /ou à de plus récentes vagues de colonisation.

Une discussion des résultats présentés dans ce manuscrit est proposée dans le **chapitre 7**. Les mammifères sont connus pour leur faible capacité de dispersion cependant nous avons montré qu'ils sont capables de coloniser de façon répétée des continents et des îles éloignées de leur continent d'origine. Les radiations de mammifères que nous observons aujourd'hui représentent les lignées qui ont survécu mais des lignées maintenant éteintes ont pu coloniser l'Amérique du Sud et Madagascar à plusieurs reprises. Nos résultats montrent que différents scénarios peuvent décrire la diversification des lignées de mammifères dans les régions nouvellement colonisées : une diversification rapide donnant naissance aux lignées encore présentes aujourd'hui (c'est le cas des rongeurs caviomorphes), une diversification différée qui a eu lieu longtemps après la période de colonisation (c'est la cas des primates platyrrhiniens), ou une diversification pas à pas, continue (c'est le cas des tenrecs de Madagascar). Il serait intéressant de confronter ces résultats avec le

registre fossile qui est, malheureusement, relativement pauvre en Amérique du Sud et non existant à Madagascar pour les lignées concernées.

CURRICULUM VITAE

I studied Medicine during two years (1994-1996) and then started Biology at the University of Science and Technology of Montpellier (1996-2000), where I got a Master degree (DEA) in Ecology and Evolution. Subsequently, I moved to Nijmegen with a one-year scholarship from the Huygens program to follow a traineeship in the Biochemistry department of the Nijmegen Centre of Molecular Life Sciences supervised by Prof. Dr. W.W. de Jong. I did my PhD degree at the Radboud University Nijmegen, Faculty of science, under the supervision of Prof. Dr. W.W. de Jong (2002-2006) with the financial support of the Netherlands Organization for Scientific Research (NWO). During my PhD I worked during four months at the Institut Municipal d'Investigació Mèdica-IMIM in Barcelona (Spain) under the supervision of Dr. M. Mar Alba. Since 2007, I work as post-doctoral fellow at the Vertebrate department of the Royal Belgian Institute of Natural Sciences (Brussels, Belgium).

LIST OF PUBLICATIONS

Poux, C., O. Madsen, W. Bergmans, E. Verheyen, and M. Vences. Molecular phylogeny and divergence times of Malagasy shrew tenrecs (*Microgale*). *In prep.*

Poux, C., O. Madsen, J. A. M. Leunissen, M. M. Alba, and W. W. de Jong. Variability of polyglutamine tracts in the mammalian androgen receptor: interrupting codons participate in the elongation process. *In prep.*

Poux C. (2009). Tenrecs and golden moles (Afrosoricida). *In* Timetree of life, S. B. Hedges, and S. Kumar, eds. Pp 479-481. Oxford University Press, New York.

Poux C., O. Madsen, J. Glos, W.W. de Jong, and M. Vences. (2008). Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses. *BMC Evol. Biol.* 8: 102.

Poux C., P. Chevret, D. Huchon, W.W. de Jong, and E.J.P. Douzery. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.* 55(2): 228-244.

Poux C., O. Madsen, E. Marquard, D.R. Vieites, W.W. de Jong, and M. Vences. (2005). Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst. Biol.* 54: 719-730.

Poux C., and E.J.P. Douzery. (2004). Primate phylogeny, evolutionary rate variations, and divergence dates: a contribution from the nuclear gene IRBP. *Am. J. Phys. Anthropol.* 124: 1-16.

Facon B., J.-P. Pointier, M. Glaubrecht, C. Poux, P. Jarne, and P. David (2003). A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. *Mol. Ecol.* 12: 3027-3039.

de Jong W.W., M.A.M. van Dijk, C. Poux, T. van Rheede, G. Kappe, and O. Madsen. (2003). Indels in protein-coding sequences of Euarchontoglires constrain the rooting of the Eutherian tree. *Mol. Phylogenet. Evol.* 28: 328-340.

Poux C., T. van Rheede, O. Madsen, and W.W. de Jong. (2002). Sequence gaps join mice and men: phylogenetic evidence from deletions in two proteins. *Mol. Biol. Evol.* 19: 2035-2037.

ACKNOWLEDGMENTS

Now that I am reaching the end of this thesis my mind goes to those who shared my Nijmegse life. First, thank all of you who, without knowing, prepared my stay in Nijmegen.

Thanks Filipe my very first non-French friend. Speaking about Portugal, its culture, its music, its football, you gave me the taste for traveling and world music. These tastes never faded away since. Without you I would have probably not even thought of going abroad.

Thank you Emmanuel for making molecular phylogenetics so attractive during your courses, without you I would have probably done oceanographical studies, I would have swam with dolphins and wheals or I would have spend years in Tahiti trying to save the coral reef. Instead I had the rare pleasure of spending my life reading chromatographs. I wish you could have been my co-supervisor but the administrative burden stopped us very fast from starting the procedure.

Thank you Frédéric for all these hours of shared work and laugh, for your knowledge, for your humor, for your kindness. Thanks Dorothée and Pascale for your help.

I would like to warmly thank you Wilfried; not everyone gets the chance of being welcomed at the station by his supervisor with a bouquet of flowers (they were yellow). At that moment I thought “ils sont sympa ces hollandais, pas si froids que ça ces nordiques!!!” Of course I had never really traveled before, I did not know what to expect, even my grandmother asked me, “Ils parlent quoi comme langue là-bas?” and I said “Hollandais mais ça semble complètement imprononçable, ça ressemble à de l’Allemand en pire, c’est dire!” Thus far it was perfect; exactly what I needed. Wilfried you are great, and all PhD students should be jealous of me. You have been so patient, always supportive but without putting pressure on me. I have always been independent: you guided me on invisible tracks. First you trusted me that I would learn English (one day but when?) and you wrote a letter stating I was speaking English very well even though I had the level of a Spanish cow (French expression).

Then you trusted me that I could finish this thesis (one day but when?) and I really hope you are proud of this piece of work. I have learned a lot with you and not only about science, I completed my course on “how to be a good person”. Thanks for your kindness and for giving me a chance.

Thank you Ole, you have been a perfect thesis partner. Working with you was so nice. I really liked your soft guidance, which gave me the feeling of freedom and safety. I will always remember our trip, crossing California in an automatic car, in the mountains at night, half asleep. You were not even annoyed each time I was breaking so softly that our heads gently flirted with the windscreen. With you even silence is nice, which is something rare and valuable in Holland.

Thank you Laura and Sergio for the time we spent deconstructing and reconstructing the world around a glass of wine.

Thanks to you my colleagues, you were nice and helpful.

Thank you Miguel for bringing the Malagasy mammals to the lab, it gave a new turn to my thesis. I was fighting with polyQs and this new project pushed up my motivation.

Erik you wanted to hire a post-doc and you got a PhD student. Thank you for giving me the time to finish my thesis; I am happy you can be part of the jury today.

All my apologies to Jack, Wilfried and Mar for not finalizing the primary focus of my thesis: the evolution of PolyQ tracks. The data are there and I still want to turn them into articles. My own rate of evolution is extremely slow and I don’t have a good relationship with deadlines but I usually reach my goals so I do believe there are good chances that this work will get published (one day but when?).

Thank you Mar for welcoming me into your lab, for making me discover Barcelona (and Ojos de brujo) and for the interesting conversations, you have been very kind.

Thanks Fernando, Elena and Carola, you have turned me into a real Catalan chica, thank you very much for welcoming me in your life so easily and warmly, I have learned so much!

Thank you all, my friends, my international family from Nijmegen, you have made my life in Nijmegen unforgettable. First Conrad, Wady and Rick. I met you the very first day I arrived in Vossendijk without any luggage, as it was lost by Air France. Thanks for having made my life so easy right away. Thanks to Xavier, Giorgia, Claudio, Vassilis, Erikur, Brecht, Emanuele, Lalo, Sacha, Salva, Marta, Anna, Catalin, Stéphanie, Arnaud, Nola, Qing, Arthur, the Spanish and the Italian communities. Writing these lines, all the good moments we have shared together come to my mind and we have shared a lot. I miss these moments and even though we are now spread all over the world, I hope our children will be able to know each other.

Merci Anna, Marie et Claire même si nous ne nous voyons pas très souvent nous cheminons ensemble.

Papa, maman merci beaucoup de m'avoir donné assez confiance en moi pour aller voir ailleurs si l'herbe était plus verte, de m'avoir toujours soutenue tout en me laissant libre de mes choix, c'est un exercice délicat, mais vous avez été d'excellents équilibristes. Merci mes petites sœurs, même quand vous êtes loin je vous sens très proche.

Αγαπες μου σας ευχαριστω. Το διδακτορικο αυτο με εφερε κοντα σας και γραψατε το καλυτερο κεφαλαιο του για μενα.